PRINCIPII

DI

[S]TATISTICA METODOLOGICA

DEL

Prof. RODOLFO BENINI

UNIONE
TIPOGRAFICO-EDITRICE TORINESE
TORINO
28 — Corso Raffaello — 28
MILANO — ROMA — NAPOLI
1906

INDICE

PARTE GENERALE

Capo Primo. — **Nozioni introduttive.**

Capo Secondo. — **La Statistica come ramo della Logica.**

Capo Terzo. — **Partizione della materia.**

PARTE SPECIALE

Libro I. — La Statistica come forma di osservazione.

Capo Primo. — La rilevazione dei dati e la formazione delle tabelle primitive.

TITOLO I. — *La rilevazione dei dati.*

TITOLO II. — *La formazione delle tabelle primitive.*

Capo Secondo. — Critica e comparazione dei dati primitivi.

Capo Terzo. — Procedimenti semplificativi e rappresentativi delle quantità rilevate.

TITOLO I. — *Procedimenti aritmetici.*

A) Teoria delle medie.

B) Interpolazione per mezzo di funzioni trigonometriche.

C) Critica dei procedimenti d'interpolazione.

D) Procedimenti interpolatorii per alcune seriazioni particolari.

E) Dell'interpolazione nel caso di osservazioni mediate.

F) Teoria delle correlazioni.

CAPO QUARTO. — **Elementi di calcolo combinatorio e di calcolo delle probabilità.**

TITOLO I. — *Elementi di calcolo combinatorio.*

TITOLO II. — *Elementi di calcolo delle probabilità.*

A) Teoremi fondamentali.

CAPO SETTIMO. — **Comparazione e critica dei dati elaborati.**

APPENDICE AL LIBRO I. — **Del Censimento.**

(In particolare dell'ultimo censimento italiano).

LIBRO II. — **La Statistica come forma di induzione.**

CAPO PRIMO. — **I sègni della presenza e del modo di operare delle cause.**

CAPO SECONDO. — **I metodi di induzione sperimentale e loro applicabilità ai fatti collettivi.**

Capo Terzo. — **L'ufficio delle ipotesi nella Statistica.**

Appendice al Libro II. — **Cenno storico intorno alla Statistica.**

PARTE GENERALE

CAPO PRIMO

Nozioni introduttive.

Sommario: § 1. Definizione della Statistica; fenomeni collettivi. — § 2. Scopo ed importanza dell'indagine numerica. — § 3. Operazioni caratteristiche: enumerazione, classificazione e graduazione, disposizione in ordine cronologico. — § 4. Il *tipico* nella varietà dei casi. — § 5. Il *costante* nella variabilità. — § 6. Il *probabile* nell'accidentalità. — § 7. Importanza del principio delle variazioni. — § 8. Concetto di causa. — § 9. Nozioni empiriche semplici, nozioni empiriche di uniformità di natura, leggi e principii. — § 10. Concetti di rapporto, correlazione, interdipendenza, funzione e limite.

§ **1.** Col nome di « Statistica » intendiamo una forma di osservazione e di induzione appropriata allo studio quantitativo dei fenomeni, che si presentano come pluralità o masse di casi, suscettive di variare senza regola assegnabile a tutto rigore.

Tali fenomeni diconsi *collettivi*. Se ne incontrano in ogni campo dello scibile: in Demografia come in Zoologia e Botanica, in Meteorologia come nelle scienze mediche, in Fisica come in Economia politica. Perfino nelle analisi della lingua e dello stile degli scrittori abbiamo applicazioni, sia pure rudimentali, dei metodi statistici.

La popolazione di un paese, aggregato mutevole di individui differenti per sesso, età, professione, religione, ecc., è un fatto o fenomeno collettivo per eccellenza; gli individui sono i casi singoli di tal fatto o fenomeno, che ci si presenta come massa indefinitamente variabile. Da un anno all'altro, da un giorno all'altro o dall'uno all'altro paese, la popolazione non sarà più la stessa per numero, per composizione, per modo di distribuzione nel territorio, ecc.; i rapporti numerici delle varie classi, pur conservandosi somiglianti a quelli rilevati prima o rilevati altrove, non saranno mai a tutto rigore identici; l'identità, quando fosse, dovrebbesi ritenere puramente accidentale.

Registrate la temperatura in una data località, ogni ora del giorno, per un intero anno. Ecco di nuovo una pluralità o massa, indefinitamente variabile, di casi. Potete esser certi che un qualunque altro anno

di osservazione darà una serie somigliante, ma non identica per il valore e per l'ordine di successione delle temperature registrate.

Al bersaglio un tiratore fa parecchie centinaia di colpi. La distribuzione dei colpi attorno al centro di mira risulterà forse molto simile da centuria a centuria, non però esattamente la stessa. Qui pure abbiamo una massa di casi indefinitamente suscettiva di variare; indefinitamente, diciamo, cioè senza legge assegnabile a tutto rigore, ma solo, come vedremo, secondo una legge approssimata.

Su 222 esemplari del *Ranunculus bulbosus*, il De Vries ne ha trovati 133 con 5 pétali, 55 con 6, 23 con 7, 7 con 8, 2 con 9 e 2 con 10. Altre collezioni gli avrebbero date proporzioni probabilmente somiglianti, non identiche, salvo per mero caso. Siamo anche qui in campo di variabilità non vincolata a regole assolute.

Fenomeni di tal genere non si possono confondere con quelle pluralità o masse di casi che sono fisse o variano secondo una regola fissa. Le montagne che oltrepassano i 4000 m. d'altezza e quelle comprese fra 3000 e 4000, fra 2000 e 3000, ecc., nella catena delle Alpi, costituiscono una pluralità di casi, ma non variabile. Il mese è una pluralità di giorni: esso comprende or 4, or 5 giorni di un dato nome, poniamo, domeniche; ma non per cause accidentali, bensì secondo una norma determinata. Un prontuario di distanze chilometriche delle stazioni ferroviarie di un paese non è una statistica. Vi hanno però fatti, che in grande siamo soliti trattare come quantità fisse, nei dettagli invece come quantità variabili. Il delta del Po si accresce ogni anno per sedimenti di materiali trasportati; gli ettari di terreno guadagnati volta per volta sul mare saranno oggetto di osservazione statistica, come una qualunque pluralità variabile di casi. Invece l'intero bacino del Po, rispetto al quale son trascurabili le piccolissime variazioni annue della foce, una volta ben determinato e misurato, vuolsi trattare come una quantità fissa.

La categoria dei fatti collettivi si allarga via via che trionfa nelle scienze d'osservazione la tendenza a sostituire termini quantitativi, specifici, esatti, ai termini qualitativi, generici e d'interpretazione oscillante quanto gli stati d'animo degli individui. I varii modi d'essere delle cose sono percepiti anzitutto come *qualità* e *quantità*, come *coesistenza* e *successione*. L'universalità degli aspetti qualitativo e quantitativo implica la sostituibilità dei termini corrispondenti (1). Così l'intensità del suono si lascia esprimere in numero di vibrazioni, il profitto degli scolari in punti di merito, la forza del vino in gradi di alcool, la gravità, la fluidità, il colorito del metro latino dalla proporzione dei dattili e degli spondei, ecc. Similmente gli aggettivi « frequente », « raro », « abbondante », « scarso » escono dalla loro indeterminatezza grazie

(1) L. Rameri, *Elementi di Statistica*, pag. 8-9, Torino, Unione Tip.-Editrice Torinese, 1898.

all'enumerazione dei casi e se le collezioni o pluralità così ottenute hanno per ulteriore caratteristica la variabilità nel tempo e nello spazio, senza legge rigorosamente definita, la competenza a trattarne dal punto di vista quantitativo spetta senza contrasto alla statistica.

La variabilità nelle condizioni suddette non si chiarisce se non come effetto di un intreccio complicato, e alla sua volta mutabile, di cause. Nei colpi al bersaglio, per esempio, la struttura dell'arma, l'intenzione del tiratore di dirigere il colpo al centro, la fissità del punto di mira, la resistenza dello strato d'aria, ecc. sono cause costanti; sarebbero invece cause variabili, irregolari o accidentali un tremito del braccio, una momentanea distrazione, per cui si lascia partire il colpo prima che sia assicurata la giusta direzione, un soffio d'aria che devia la palla, ecc. Così la temperatura di un luogo ha per causa variabile regolare la posizione della terra rispetto al sole; per costanti: l'altitudine, la prossimità o lontananza da grandi masse d'acqua od altro; per cause accidentali: la presenza o assenza di nubi, di venti, e così via.

Questo intreccio di cause può costituire e costituisce infatti una difficoltà notevole dell'analisi, ma non è motivo che tolga interesse alla rilevazione numerica dei casi o che debba dissuadere *a priori* da ogni tentativo diretto a distinguere, sia pure in via di semplice approssimazione, la quantità di effetto che va attribuita ad alcune cause specificamente assegnabili, da quella che spetta in blocco ad altre cause, fino alle quali non giungono i nostri mezzi d'analisi.

La solidarietà dei diversi rami del sapere implica che si facciano convergere i diversi modi d'indagine alla conoscenza più completa possibile dei fatti. Certo altra cosa è sapere con quale frequenza la grandine visita questa o quella zona o quanti maschi nascono ogni 100 femmine o in quali proporzioni una data malattia colpisce gli individui della tale o tal'altra età, della tale o tal'altra professione, ecc.; ed altra cosa è accertare il processo di formazione della grandine, l'origine dei sessi o la causa specifica di quella malattia; nondimeno è chiaro che l'analisi statistica varrà nei casi supposti a mettere in risalto le circostanze in cui più frequentemente si avvera il fenomeno e a circoscrivere così l'àmbito in cui la ricerca delle cause è più promettente di risultati positivi.

§ 2. In termini generali, noi crediamo di poter così esprimere il còmpito della Statistica: « sceverare nei fenomeni collettivi ciò che vi ha di tipico nella varietà dei casi, di costante nella variabilità, di più probabile nell'apparente accidentalità e decomporre, fino al limite che la natura del metodo consente, il sistema di cause o forze, di cui essi fenomeni sono la risultante ».

Or bene, i tentativi di questa specie, in varie direzioni delle nostre conoscenze, sono dei più incoraggianti. A torto si crederebbe che il numero, come forma instabile di manifestazione dei fenomeni in parola,

non possa darci che un loro aspetto affatto secondario, irriducibile a leggi e irrilevante per la soluzione del problema delle cause: chè al contrario quasi non c'è ordine di fatti collettivi, in cui l'investigazione statistica non sia riuscita a cogliere un contenuto essenziale, una norma valida per la massa dei casi, se non pei casi singoli, un criterio razionale di classificazione, una correlazione di elementi esprimibile con formola approssimata, ma semplice. Gli scettici della statistica non hanno ragione se non di fronte ai dilettanti più superficiali di questa disciplina. I fenomeni collettivi si debbono studiare sotto l'aspetto del numero, come i non collettivi si studiano dal punto di vista della figura geometrica, della situazione, del movimento, delle proprietà fisiche o chimiche, del valore economico, ecc. Ad ognuna di queste forme di manifestazione corrisponde una scienza o un capitolo di scienza astratta o concreta. Così, la figura è studiata in astratto dalla geometria, in concreto dall'architettura, scoltura, ecc.; la forza e il movimento, in astratto, dalla meccanica razionale, in concreto, dalla meccanica celeste e da quella applicata alle arti ed industrie; il valore, in astratto, dalla economia politica, ecc. *A priori* nessuno vorrà qualificare una di dette forme di manifestazione come secondaria rispetto alle altre; se una gerarchia di dignità si dovesse stabilire, il criterio più accettabile sarebbe quello del grado relativo di perfezione raggiunto dalle competenti discipline; ma anche in tale ipotesi, quella particolar forma di manifestarsi dei fenomeni collettivi, che è la massa numerica variabile, potrebbe vantare d'aver dato origine, nello studio astratto, ad un ramo importante della Logica, l'*induzione statistica;* nelle applicazioni, alla demografia, a gran parte della meteorologia e a non pochi interessanti capitoli d'altre scienze naturali o sociali.

§ 3. Le operazioni caratteristiche che si fanno a riguardo dei fenomeni collettivi sono: l'enumerazione dei casi, la loro classificazione e graduazione, la loro disposizione in serie cronologiche. Di esse, dal punto di vista tecnico e logico, si parlerà più ampiamente nella parte speciale; qui in sede di nozioni generali introduttive ci limiteremo alla definizione e a qualche chiarimento.

L'*enumerazione* si risolve nella addizione aritmetica dei casi considerati ciascuno come una unità equivalente a qualsiasi altra unità della stessa classe.

La *classificazione* è l'operazione per la quale separiamo l'uno dall'altro gruppi di casi che, pur avendo le caratteristiche essenziali comuni della specie, differiscono tuttavia per elementi secondari *discontinui*, cioè tali che non ammettono il passaggio dall'uno all'altro per gradazioni infinitesimali. Così se distinguo gli individui di una popolazione secondo il sesso, lo stato civile, la religione, ecc., la classificazione si dirà per elementi discontinui, non essendoci termini di graduale passaggio da un sesso all'altro, dalla condizione di celibe a quella

di ammogliato o di vedovo, dal culto cattolico all'israelitico o all'evangelico e via dicendo. E ancora, se una Banca emette biglietti con tagli fissi, mettiamo, da 50 lire, da 100, da 500 e da 1000, la classificazione corrispondente si dirà ancora per elementi discontinui, mancando i biglietti rappresentanti gli infiniti tagli intermedi che si potrebbero concepire.

La *graduazione* invece è l'operazione per cui separiamo i casi in gruppi secondo caratteri che ammettono termini, quanti si vogliano, di transizione. Posso, ad es., graduare gli individui per età, per forza misurata al dinamometro, per reddito; i trasporti di merce secondo il peso o la distanza; le vendite di immobili secondo il valore o l'estensione, ecc., questi elementi ammettendo infinite gradazioni. Solo per comodità dell'enumerazione e della presentazione dei dati, convien procedere per intervalli più o meno comprensivi, come di anno in anno di età, di dieci in dieci chilometri di distanza, di mille in mille lire di reddito.....

Classificazioni e graduazioni si possono fare per *elementi isolati* o per *elementi combinati;* in altre parole il fenomeno vien trattato, per dirla in linguaggio matematico, o come funzione di una sola variabile o come funzione di due o più variabili. Una schiera di nati si può distinguere secondo l'età del padre soltanto o secondo l'età della madre soltanto o secondo le età combinate dei genitori; un insieme di famiglie secondo il numero dei componenti o quello delle stanze di abitazione o contemporaneamente secondo l'una e l'altra variabile. È ovvio che, quanto maggiore è il numero degli elementi combinati, tanto più *scelti* sono i gruppi che risultano dalla specializzazione.

La *disposizione in serie cronologiche* è la distinzione dei casi secondo gli intervalli di tempo in cui si sono successivamente verificati. La divisione civile del tempo in anni, mesi, settimane, giorni ed ore fa regola per la scelta degli intervalli.

Il principio di Logica, giusta il quale ciò che si afferma di un'intera classe di oggetti può ripetersi di ogni singolo oggetto o gruppo di oggetti, va inteso con discrezione nell'ordine dei fatti di competenza della Statistica. La indefinita variabilità, che dicemmo essere caratteristica della massa dei casi, si ritrova ancora e sempre come caratteristica dei gruppi scelti, in cui la massa può frazionarsi. Qui il principio in parola ha piena applicazione. Anzi un indizio decisivo della molteplicità delle cause, dalle quali dipendono i fenomeni collettivi, è questo appunto, che classificandoli e graduandoli in tutti i modi immaginabili noi ci troviamo, prima come poi, in presenza di quantità essenzialmente variabili e mai di quantità fisse o variabili secondo legge definita a rigore. Così, per esemplificare, non solo è variabile la cifra totale dei suicidii, che accadono in un paese, ma lo son pure, ed anzi lo sono maggiormente, le cifre parziali dei suicidii di maschi o di femmine e quelle ancor più parziali che si ottengono combinando l'elemento del sesso con l'altro dell'età e ricombinando questi due elementi con quello dei

luoghi (città e campagne, provincie del nord, del sud, ecc.) o dei mezzi prescelti o d'altro ancora. Invece, sotto altri aspetti, si dà una antitesi spiccata fra caratteri dei casi singoli o dei gruppi scelti e quelli del loro insieme indistinto. È ciò che ora dimostreremo.

§ **4.** Nella varietà dei casi lo statistico si propone di ricercare il *tipico*. Che cosa s'intende per « tipo »?

In alcune scienze ed arti lo si intende come « modello » che risponde ad una idea nostra di perfezione o a certe condizioni, dalle quali deriva un *maximum* di effetto estetico, un *maximum* di effetto utile od altro — armonia di linee e di proporzioni come in una statua; economicità di organizzazione come in una Banca, ecc. Oppure nel senso di « esemplare » convenzionalmente scelto, cui vengono riferiti i casi della stessa categoria (il campione internazionale del metro, i minerali scelti come tipi della graduatoria di durezza o dei sistemi di cristallizzazione e simili). In etnografia dicesi tipico il soggetto in cui si trovano associati i caratteri predominanti di una razza; così il tipo del Sardo non sarà realizzato in quell'individuo che sia soltanto basso di statura o soltanto dolicocefalo o soltanto bruno d'occhi e di capelli, ecc., ma in quello che è ad un tempo basso, dolicocefalo, bruno, ecc. Gli statistici alla lor volta considerano tipico il carattere che appartiene alla classe di maggior frequenza, alla varietà più comune fra molte varietà più o meno rare, talora anche la semplice media risultante da un gran numero di osservazioni; criterio che certo non collima con quello che identifica il tipo coll'idea di modello o di perfezione. Al Quételet fu rimproverato di non avere, nella sua concezione dell'*uomo medio*, tenuti ben distinti i due criteri.

Estendendo un po' più l'esplorazione, noi troviamo dominii dello scibile, nei quali anche un caso singolo, ben accertato, fa regola per tutti i casi della stessa classe, presenti o futuri. Il mercurio puro di qualunque provenienza ha il peso specifico di 13,6, si solidifica a 40° sotto zero, si amalgama coll'oro, ecc.; una qualunque corda, che per una certa lunghezza dà il *do*, per i due terzi esatti di quella lunghezza dà il *sol;* e così via. Qui non c'è bisogno di calcolare su grandi masse, nè ci sono gradazioni *reali* dello stesso fenomeno rappresentate da un maggior numero di casi, in confronto di altre gradazioni (1); i fatti sono tipici per sè medesimi e fra un tipo e l'altro havvi discontinuità. La legge delle proporzioni definite in chimica e quella delle proporzioni

---

(1) Possono darsi gradazioni *apparenti*, imputabili cioè a discordanze ed errori di osservazione; ma allora sono le osservazioni stesse che colle loro divergenze costituiscono una pluralità di casi trattabile mediante i metodi statistici, come qualsiasi altro fenomeno collettivo; mentre la grandezza, cui si riferiscono le osservazioni, è quella che è, e fa regola per i casi della medesima specie.

multiple significano appunto che vi hanno discontinuità o salti e non si passa per gradi infinitesimali da una combinazione all'altra. I minerali cristallizzano secondo sei sistemi e 32 tipi; ma gli innumerevoli gradi di transizione immaginabili da un tipo all'altro mancano di rappresentanza propria. In quella vece, in altre direzioni delle nostre conoscenze, massime nel mondo organico e sociale, dove la complicazione delle cause e le interferenze delle leggi pongono a seria prova lo spirito di analisi, l'individuo o il fatto singolo raramente fa regola per l'intera classe cui appartiene o la fa solo per certi caratteri generali e non per i caratteri particolari: di solito gli è dall'insieme dei casi formanti entro certi limiti un *sistema continuo di gradazioni dello stesso fenomeno*, che noi possiamo trarre l'elemento tipico (1). Ad una condizione però: che i casi discostantisi dalla media risultino tanto più rari quanto più divergenti o eccezionali. Allora la gradazione, cui corrisponde il gruppo di maggior frequenza, costituisce in certo modo il centro di gravità del sistema e dà il tipo: essa si chiarisce come l'effetto normale di cause costanti non disturbate da cause accidentali, sia perchè queste spiegano un'azione troppo debole, sia perchè si sono l'una l'altra perfettamente neutralizzate.

Stature	N° degli esaminati su 10.000	Età	N° dei vedovi che sposano vedove
1,48-1,50	66	22-25	23
1,50-1,52	121	25-28	93
1,52-1,54	187	28-31	190
1,54-1,56	500	31-34	327
1,56-1,58	751	34-37	451
1,58-1,60	945	37-40	571
1,60-1,62	1.171	40-43	645
1,62-1,64	1.244	43-46	663
1,64-1,66	1.224	46-49	637
1,66-1,68	1.093	49-52	636
1,68-1,70	873	52-55	647
1,70-1,72	651	55-58	582
1,72-1,74	435	58-61	497
1,74-1,76	276	61-64	332
1,76-1,78	165	64-67	273
1,78-1,80	88	67-70	199
1,80-1,82	47	70-73	117
ecc.		ecc.	

Il lettore, che voglia concedere un po' d'attenzione ai due esempi numerici qui riportati, l'uno sulla graduazione per stature dei nostri coscritti, l'altro sulla graduazione d'età dei vedovi che sposano vedove,

(1) Nella teoria darwiniana si dà a questo concetto di « continuità » una portata estrema, poichè si ammette la discendenza delle specie da uno o da pochi germi primordiali per via di successive graduali trasformazioni. Ipotesi, convien dirlo, recentemente impugnata dallo Schiaparelli, il quale ritorna

non può fraintenderci. Le due masse di casi gli si presentano come sistemi continui di gradazioni — fino a certi limiti, si capisce. — Il gruppo più numeroso nella serie delle stature è quello che misura da 1,62 a 1,64; nella serie delle età è quello che conta da 46 a 49 anni (1). A partire dalla statura di m. 1,63 o dall'età di 47 $^1/_2$ anni nelle sole direzioni possibili, che son due, i gruppi di soggetti osservati impiccioliscono via via tanto più rapidamente, quanto più avanziamo verso le stature giganti o retrocediamo alle nane e, rispettivamente, verso le età troppo mature o verso quelle troppo immature per seconde nozze. Grazie a questa simmetria o quasi simmetria di distribuzione e alla prevalenza numerica dei gruppi di età media e di media statura, il centro di gravità del sistema (ci si permetta il paragone dei casi di un fenomeno collettivo coi punti pesanti di un corpo) si individua, si localizza in cotesti gruppi centrali o in prossimità di essi. Questi forniscono dunque l'elemento tipico cercato, volendosi significare con ciò che la collaborazione delle molteplici cause da cui dipende la statura dei ventenni o l'età dei vedovi in seconde nozze con vedove produce in corrispondenza della misura di 1,63 o dell'età di 47 $^1/_2$ anni un *risultato massimo.*

Comunque s'intenda però il tipo in statistica, o nel senso astratto di media o in quello concreto di carattere della varietà numericamente più forte, esso non ha nulla di assoluto. Infatti, dato un complesso di casi graduabili o classificabili nei modi più diversi, è in nostro arbitrio, come già dicemmo al § 3, di raggrupparli per elementi isolati o per elementi combinati. Corrispondentemente avremo categorie più o meno scelte e tipi più o meno specifici. Immaginiamo di ripartire 10.000 neonati per gradazioni di peso, di cento in cento grammi; la gradazione che è numericamente rappresentata più d'ogni altra formerà il tipo generico « peso del neonato ». Siano ora suddivisi i soggetti d'osservazione secondo la origine loro (legittima o illegittima), secondo l'ordine di genitura (primogeniti, secondogeniti, ecc.), il modo della nascita (nati da parti semplici o invece gemelli, trigemini), l'età della madre, la condizione economica della famiglia, ecc.; i nuovi gruppi risulteranno assottigliati di numero e quelli di maggior frequenza relativa in ogni suddivisione forniranno i valori tipici di

---

all'idea della discontinuità probabile anche nei tipi organici, argomentando per analogia da quella dei tipi del regno inorganico e movendo da considerazioni matematiche, che qui non possiamo riprodurre (Vedi G. V. SCHIAPARELLI, *Studio comparativo tra le forme organiche naturali e le forme geometriche pure.* In appendice alle *Peregrinazioni antropologiche e fisiche* del dott. TITO VIGNOLI; Biblioteca Scientifico-Letteraria, Milano, Hoepli, 1898).

(1) Secondo questi dati, tratti dal *Movimento dello stato civile in Italia nel* 1878, il gruppo più numeroso parrebbe quello dai 43 ai 46 anni; ma siccome esso, che contiene l'età rotonda di 45, si è probabilmente ingrossato per erronee denunzie d'età a scapito del successivo, così crediamo di dover attribuire al gruppo 46-49 il *maximum* di frequenza e il centro di gravità della seriazione.

peso distintamente secondo l'origine, l'ordine di genitura, ecc., valori più o meno divergenti da quello calcolato sull'insieme indistinto delle nascite. Si troverà così che il peso medio dei neonati illegittimi è inferiore d'un certo tanto a quello dei legittimi; che il peso medio dei nati da madri precoci è inferiore di tant'altri grammi a quello dei nati da madri in piena fioritura d'età, e così via. Ma possiamo andar oltre e combinare due a due, tre a tre gli elementi fissi di classificazione, cioè formare gruppi scelti di neonati che siano ad un tempo illegittimi e figli di madri precoci e povere o legittimi di madri in pieno sviluppo d'età e di condizione agiata, o legittimi ma di madri pluripare, avanzate in età e di media condizione, ecc., ecc., fino ad esaurire tutte le combinazioni possibili. I gruppi così ottenuti si moltiplicheranno, assottigliandosi sempre più di contenuto e tra essi ci saranno quelli di maggior frequenza relativa, che forniranno i tipi delle rispettive categorie e sottocategorie, tipi più o meno divergenti ancora da quello generale o comune. Procedendo in questa maniera fino ai limiti, che vedremo consentiti alla specializzazione, noi potremo riprodurre molti dei valori già osservati sui soggetti singoli, i quali valori saranno tipici per rispetto ai gruppi così selezionati, mentre nella massa figuravano come casi di deviazione dal tipo. Si arriva così alla importante conclusione: *Gli elementi tipici per gruppi scelti cessano di essere tipici nella massa, nella quale si comportano piuttosto come casi di deviazione dalla norma generale.* Ed è facile convincersi che sussiste la proposizione inversa: *L'elemento, che è tipico per la massa, può figurare come caso di deviazione nei gruppi scelti di questa* (1).

Ecco una prima antitesi di caratteri tra il fenomeno collettivo considerato come un tutto e i casi singoli o gruppi scelti di casi, di cui si compone. Ora ne troveremo delle altre.

§ 5. Lo statistico si propone ancora di determinare ciò che vi ha di *costante* nella *variabilità* dei fatti collettivi, osservati nella successione del tempo. Per « costante » non vuolsi intendere l'immutabile

(1) Un tentativo ingegnoso di realizzare il concetto statistico del tipo in un campo che parrebbe aperto alle più variabili impressioni degli osservatori fu fatto dal GALTON (*Inquiries in to human faculties*, London, Macmillan, 1883) col procedimento delle « fotografie multiple o composte ». Comunemente parliamo di un « tipo di famiglia », di un « tipo del delinquente », del « tipo dell'ebreo, dell'inglese, ecc. », e gli abili caricaturisti sanno, anche esagerando certi tratti, farcelo riconoscere nel soggetto del disegno. Ciò significa come nella nostra memoria si siano fissate le impressioni di alcune caratteristiche, che son comuni agli individui di certi gruppi scelti e che li differenziano da quelli di altri gruppi. Ora il Galton, esponendo successivamente, in qualunque ordine, alla lastra fotografica i ritratti di parecchie persone (es., i membri di una stessa famiglia), avendo cura che fossero previamente presi nella stessa

in modo assoluto, ma il relativamente costante, il variabile regolare, o ciò che procede in funzione piuttosto semplice del tempo. Come il concetto di « tipo » non ha nulla di assoluto, così la separazione della parte variabile irregolare da quella costante o regolarmente variabile in una serie è un problema che si risolve solo partendo da postulati più o meno arbitrari. Di variabilità irregolari se ne possono immaginare molte per una medesima serie statistica, come si possono dare tipi più o meno specifici in una massa frazionata a piacere in gruppi scelti.

Abbiasi, ad esempio, la serie dei matrimoni contratti in Italia nel trentennio 1872-1901.

Anni	Matrimoni	Variazioni da un anno all'altro	Anni	Matrimoni	Variazioni da un anno all'altro
1872	202.361	.....	1887	235.629	+ 2.319
1873	214.906	+ 12.545	1888	236.883	+ 1.254
1874	207.997	— 6.909	1889	230.451	— 6.432
1875	230.486	+ 22.489	1890	221.972	— 8.479
1876	225.453	— 5.033	1891	227.656	+ 5.684
1877	214.972	— 10.481	1892	228.572	+ 916
1878	199.885	— 15.087	1893	228.103	— 569
1879	213.096	+ 13.211	1894	231.581	+ 3.478
1880	196.738	— 16.358	1895	228.152	— 3.429
1881	230.143	+ 33.405	1896	222.603	— 5.549
1882	224.041	— 6.102	1897	229.041	+ 6.438
1883	231.945	+ 7.904	1898	219.597	— 9.444
1884	239.513	+ 7.568	1899	235.665	+ 16.068
1885	233.931	— 5.582	1900	232.631	— 3.034
1886	233.310	— 621	1901	234.819	+ 2.188

Facendo la media annua del primo quindicennio (= 219.918) e quella del secondo (= 229.557) si trova che la seconda supera la prima; è ovvio concludere che il movimento della nuzialità nel nostro paese, a prescindere dalle piccole oscillazioni annuali, è *in generale* ascendente. Graficamente si potrebbe rappresentarlo con una retta che, passando per gli estremi di due ordinate proporzionali alle medie in

posizione di fronte o di profilo, nelle stesse dimensioni, colla luce dalla stessa parte, riuscì a mettere in evidenza nella fotografia composta risultante i caratteri comuni, attorno ai quali compaiono con lievi sfumature e incerte linee i caratteri divergenti dei singoli soggetti. Se rassomiglianza esiste, apprezzabile già all'occhio, per es., tra due fratelli, l'immagine media ottenuta dalla fotografia composta è tale, che si può attribuirla all'uno o all'altro dei due, quasi indifferentemente. Colla solita genialità il Galton suggerì ancora questo metodo come l'unico che possa in certi casi mettere d'accordo le fisionomie più o meno discordanti attribuite ad uno stesso personaggio storico nelle monete, medaglie, pitture. Così da sei diverse medaglie di Alessandro Magno egli ricavò una figura unica, che nell'incertezza di scelta tra le singole figure può ritenersi la più verosimile.

questione, tende ad elevarsi sempre più sull'asse dei tempi. La parte variabile regolare sarà così indicata dal movimento rettilineo (1). Facciansi ora invece le medie del primo, secondo e terzo decennio; esse sono rispettivamente: 213.604; 231.533 e 229.076. La differenza tra la prima e la seconda è notevole e positiva; quella tra la seconda e la terza, piccola ma negativa. La nuzialità è cresciuta da principio e rapidamente; indi è lentamente declinata. L'andamento *generale* della serie, a prescindere sempre dalle oscillazioni annue, non risulterà più rettilineo: per gli estremi di tre ordinate proporzionali alle tre medie in questione si può far passare una parabola ordinaria; questa curva, e non più la retta di poco fa, rappresenterà graficamente la parte variabile regolare del fenomeno; tutto ciò che sta fuori di questo tracciato è o si presume essere elemento irregolare o accidentale. Senonchè nulla vieta d'andar oltre e calcolare le medie dei quattro periodi di 7 anni e mezzo, in cui si può spezzare la serie, o dei cinque sessennii o dei sei quinquennii e, graficamente, far passare per i punti che le rappresentano delle curve di 3°, 4°, 5° grado, le quali ancora figureranno l'andamento « generale » della serie, a prescindere dalle anomalie dei singoli anni. Evidentemente però il carattere di generalità svanisce man mano che si moltiplicano i punti, per cui obblighiamo la curva teorica a passare. Quand'è che avremo una generalità « sufficiente », cioè nè troppo vaga, nè inquinata da elementi accidentali? Noi non possiamo dirlo con criteri assoluti. Il problema, che è dei più ardui e interessanti della statistica metodologica, sarà ampiamente trattato nel capitolo degli *Aspetti arbitrari dell'interpolazione*. Qui basti di avervi accennato.

Il concetto di « costanza nel tempo » richiama quello di cause permanenti o variabili, regolari e non intermittenti o saltuarie: quindi involge ancora l'idea di « continuità ». *Continuità* è quel modo di manifestarsi dei fenomeni collettivi, per cui essi (se si osservano in masse abbastanza grandi) passano ininterrottamente e per gradi da certe condizioni di frequenza a certe altre, col trascorrere del tempo. *Discontinuità* è il modo opposto.

Vi ha continuità o assenza di bruschi salti, per es., nella curva di sviluppo dell'individuo dalla nascita in avanti. Una classe di coetanei non si estingue tutta d'un tratto, ma gradatamente. La fecondità della donna diminuisce d'anno in anno fino a scomparire del tutto. ecc.

La discontinuità suppone intervento improvviso di *cause nuove*, a punti determinati della serie. Una guerra di tariffe tra due paesi

---

(1) I lettori, ai quali non fossero già per altra via famigliari i metodi grafici, quanto occorre per intendere alcune espressioni speciali qui usate, possono consultare anticipatamente i primi paragrafi del capitolo sui *Procedimenti geometrici e grafici* nella Parte speciale.

riduce istantaneamente le loro reciproche importazioni di merci. La serie dell' emigrazione inglese presenta brusche variazioni all'epoca della scoperta delle miniere aurifere d'Australia e di California. Il numero degli elettori politici fu più che triplicato nel 1882 per la riforma introdotta in Italia colla legge 24 settembre di quell'anno.

Perfettamente realizzata la continuità nel tempo non si può mai dire; le stesse cause che diciamo accidentali agiscono appunto volta per volta, quando non si neutralizzano in modo completo, come *cause nuove*, sopravvenienti a perturbare l'andamento regolare del fenomeno.

Anche qui si palesa un'antitesi frequente di caratteri tra i singoli casi o gruppi scelti di casi e il loro insieme.

Ci sono funzioni o manifestazioni discontinue presso l'individuo, che diventano continue nella massa, purchè questa sia composta di un sufficiente numero di individui in condizioni diverse l'uno dall'altro. Tizio passa una o due volte al giorno per una certa via; qualche volta non ci passa affatto: ecco una funzione discontinua. Pure per quella via passa sempre gente, perchè la massa si compone di soggetti in condizioni svariatissime per rispetto al bisogno o all'opportunità di percorrere quella via: ecco una funzione continua. Similmente gli individui A, B, C non comperano caffè o zucchero più del solito, se il prezzo di tali derrate diminuisce solo del 10 o del 15 %: ne comprerebbero di più se il prezzo scendesse almeno di 20 %: ecco una funzione discontinua. Ma la massa è così varia nella sua composizione, che ad ogni benchè piccolo ribasso del prezzo corrisponde un aumento del consumo, ad ogni rincaro una restrizione della domanda (1). Che cosa di più discontinuo del nascere e del morire, per l'individuo? Si nasce e si muore una volta sola; ma in seno ad una numerosa popolazione avvengono nascite e morti quasi ad ogni istante.

Gli è chiaro però che non sussiste la reciproca: un fenomeno, che si presenta come funzione discontinua nella massa, non può essere continuo presso l'individuo o gruppi scelti di individui. Ogni dimostrazione ci sembra superflua.

Un caso particolare della « costanza » è la *periodicità*, ossia il ripetersi dei massimi e dei minimi di un fenomeno ad intervalli di tempo piuttosto regolari. In meteorologia, la temperatura, la pressione barometrica, la quantità di elettricità contenuta nell'atmosfera, ecc., presentano caratteri di periodicità; nella fisica terrestre, le maree, il livello dei fiumi, ecc.; in demografia, i matrimoni, le nascite, le morti, l' emigrazione temporanea; in economia, il movimento ferroviario,

---

(1) Gli economisti matematici trattano spesso gli atti dell'*Homo oeconomicus* come funzioni continue, supponendo in lui, per comodità di ragionamento deduttivo, una sensibilità perfetta alle variazioni dell'ambiente economico. La ipotesi è senza dubbio irreale, se si guarda ai singoli individui concreti, ma si avvicina alla realtà se si guarda all'insieme, alla massa.

quello dei depositi e sconti presso le banche, molte forme di consumi. Giuseppe Ferrari ne' suoi *Periodi politici* ha introdotto, con tentativo discutibile, l'idea di periodicità negli avvenimenti della storia civile.

Il succedersi delle stagioni, il succedersi del giorno e della notte sono le cause naturali della periodicità di un gran numero di fatti collettivi. Anche le divisioni artificiali o convenzionali del tempo hanno una certa importanza. Nei matrimoni, ad es., si osserva una ben distinta periodicità settimanale, con un minimo nel venerdì e un massimo generalmente in domenica; essi hanno al tempo stesso una periodicità annua segnalata da due massimi e due minimi speciali nei mesi precedenti o coincidenti colla Quaresima o coll'Avvento. Presso le nostre casse postali di risparmio fu notato che i depositi superano i rimborsi nella prima quindicina d'ogni mese; sono invece inferiori ai rimborsi nella seconda, della qual cosa non è difficile trovare la spiegazione; ed oltre a ciò presentano due massimi eccezionali nel gennaio e nel luglio, in coincidenza delle scadenze della rendita pubblica, che si capitalizza a risparmio. Sicchè possiamo avere in una serie l'intreccio di periodicità diverse (1).

Anche qui campeggia il fatto che *una funzione periodica per l'individuo singolo o per gruppi scelti d'individui, può perdere o modificare questo carattere di periodicità nella massa.* Basta riflettere ai casi in cui al massimo della funzione per un individuo o per un gruppo corrisponde il minimo della funzione per un altro individuo o un altro gruppo, e viceversa. Avviene allora una specie di compensazione ripetuta, che tende ad imprimere al fenomeno, guardato nella sua totalità, un andamento uniforme, senza ondulazioni caratteristiche e ricorrenti. È a deplorarsi che in questo campo, in cui c'è molto da mietere, non si abbiano indagini nè teoriche, nè statistiche. Chi scrive ha tentato altra volta in tema di circolazione monetaria di analizzare (ma da un punto di vista teorico soltanto) la formazione delle giacenze o disponibilità note sotto il nome di *hoards*, la quale si chiarisce appunto come risultato di una compensazione di periodicità diverse. L'osservazione comune infatti insegna che nei mesi in cui, ad esempio, gli agricoltori vendono i raccolti e portano il denaro disponibile a deposito presso le banche, altre categorie di produttori hanno invece un *minimum* di vendite e un *maximum* di acquisti (per rifornirsi di materie prime, ecc.) e ritirano dagli istituti di credito le somme già loro affidate in altre epoche dell'anno; sicchè le giacenze o disponibilità presso gli istituti, che hanno una clientela variamente assortita, tendono a mantenersi ad un livello piuttosto costante.

---

(1) Uno studio ancora da farsi e pel quale non difettano gli elementi sparsi oggi in una grande quantità di statistiche ufficiali, sarebbe appunto quello della connessione del maggior numero possibile di fenomeni collettivi, naturali o sociali, dal punto di vista della periodicità.

Perchè un fenomeno periodico presso i singoli individui conservi il carattere di periodicità nella massa, occorre che sia *tipico;* occorre cioè che la varietà delle situazioni individuali non sia tale da favorire una ripartizione di casi uniforme nel tempo, ma costituisca, entro certi limiti, un sistema continuo, nel quale i casi divergenti dalla media in qualunque senso risultino tanto più rari, quanto più divergenti. Le nostre quotidiane abitudini ci forniscono svariati esempi. Il dormire è un fenomeno periodico per i singoli individui, ma non cessa di essere tale anche per la massa, appunto perchè, nonostante la varietà delle disposizioni individuali, che arriva a coloro i quali fanno della notte giorno e viceversa, le ore della notte sono le preferite dalla gran maggioranza. Nelle trattorie ci sono ore tipiche ossia di maggior frequenza degli avventori, il cui numero si assottiglia invece, entro certi limiti, quanto più si anticipa o si posticipa su tali ore. Però se la varietà delle situazioni individuali non sopprime in siffatte circostanze il carattere di periodicità nella massa, tende tuttavia a convertire i casi singoli discontinui in fenomeno collettivo continuo.

Considerando le cose da un punto di vista più generale, diremo che qualunque sia il movimento di un fenomeno collettivo osservato nella successione del tempo, esso può essere la risultante di movimenti molto diversi nei gruppi scelti o nei casi individuali, di cui la massa si compone. Un andamento ascensionale della massa è compatibile con quello discendente di una parte di essa. Così, il consumo del tabacco in genere tende in Italia a crescere; ma, nella specie, il consumo del tabacco da fiuto da vent'anni in qua seguita a diminuire. Un movimento piuttosto rettilineo dell'insieme dei casi (cioè il loro crescere o scemare di quantità eguali o pressochè eguali ad ogni unità di tempo) potrebbe risultare dalla compensazione di due inversi movimenti parabolici nei principali gruppi scelti, in cui si può bipartire la massa. Sicchè, in ultima analisi, la specializzazione dei dati, fino al limite utile al quale può essere portata, appare l'unico mezzo per accertare l'intreccio e l'interferenza delle leggi che governano il fatto in esame.

§ **6.** Altro immediato obbiettivo è la ricerca di ciò che è ***più probabile*** nell'apparente ***accidentalità*** e ***indipendenza*** dei casi.

Della « probabilità » havvi un concetto comune, soggettivo, e un concetto matematico ed oggettivo. Il primo si presenta, quando le circostanze favorevoli o contrarie al verificarsi di un dato avvenimento non sono suscettive di rigorosa determinazione e possono perciò essere diversamente apprezzate dai diversi individui. Due persone, che aspettano un comune amico ad un appuntamento, possono attribuire un diverso grado di probabilità alla sua venuta, secondo l'opinione che ciascuna si è formata della puntualità dell'amico, del suo inte-

resse a non mancare al convegno, degli impegni che forse lo trattengono altrove, ecc. Quando invece i casi favorevoli ed i contrari a un dato avvenimento sono *numerabili* con esattezza e v'è ragion di credere che siano tutti *possibili in egual grado*, entriamo nel campo della probabilità oggettiva o matematica. Così la probabilità di fare il punto 5 gettando un dado diciamo essere pari ad $^1/_6$, perchè 6 sono le faccie del dado con egual possibilità di sortita (se almeno il dado è ben costrutto) e 1 sola è la faccia che porta impresso il punto 5. Similmente la probabilità di estrarre un « fante » da un mazzo di 40 carte è eguale a $^4/_{40} = ^1/_{10}$, perchè 40 sono le carte con egual possibilità di sortita e 4 sono i fanti, cioè i casi favorevoli all'aspettazione. Qui non si può dare ragionevolmente un giudizio diverso da individuo ad individuo; il credere alle carte « buone » e ai numeri « simpatici » è superstizione che si elide o si bilancia colla superstizione opposta d'altri giuocatori. La probabilità matematica è dunque il rapporto tra il numero dei casi favorevoli ad un certo evento aspettato e quello dei casi possibili, rapporto affatto indipendente dalla nostra fiducia.

Però nei fenomeni collettivi, che certo dipendono da un intreccio complicato e mutevole di cause in parte costanti o variabili regolari, in parte variabili irregolari, tutto ciò che possiam fare per calcolare la probabilità di un evento è di numerare i casi avvenuti in passato e metterli in rapporto con quelli che per estrema ipotesi avrebbero potuto verificarsi, supposto poi costante tale rapporto e applicabile ai casi futuri. Su 1000 nubili ventenni, che tutte per estrema ipotesi avrebbero potuto maritarsi nel corso del ventunesimo anno d'età, solo 95 in media, giusta l'osservazione del passato, sono effettivamente passate a nozze nel corso dell'anno. 95 $^0/_{00}$ è la probabilità applicabile alle ragazze ventenni dell'oggi o del domani. In siffatto modo il concetto statistico di probabilità viene ad avere un posto di mezzo tra quello comune e quello matematico. Si avvicina al concetto matematico, in quanto si definisce come un rapporto tra casi avvenuti e casi che avrebbero potuto avvenire, rapporto oggettivo e non sottomesso al capriccioso apprezzamento degli individui; ma si avvicina anche al concetto comune, in quanto implica: 1° la supposizione che detto rapporto si conserverà invariato in futuro, idea alla quale individui diversi possono attribuire un diverso valore; 2° la supposizione che i singoli casi siano possibili in egual grado, mentre in realtà nol sono. Questo secondo punto è di speciale importanza e merita schiarimento. Le carte di un mazzo, su cui metto le mani a caso, magari ad occhi chiusi, possono ben considerarsi come aventi la stessa possibilità di sortita; ma le nubili disponibili sono invece esposte ad una *scelta* vera e propria, sulla quale influiscono le speciali condizioni di ciascuna. Le già fidanzate, con promessa di matrimonio a breve scadenza, sono in una posizione migliore, *caeteris paribus*, delle fidanzate a scadenza

più lontana o indeterminata: le une e le altre in posizione migliore delle non fidanzate: e tra queste le deformi, le malate di malattie che non lasciano speranze, le votate per propria elezione alla verginità, si trovano nelle condizioni men favorevoli. Tuttavia, nella massa, noi operiamo idealmente un conguaglio e supponiamo i casi tutti possibili in egual grado, senza di che riuscirebbe insolubile il quesito di formulare per i fenomeni di massa una probabilità determinata, mentre nei casi singoli, di cui quella si compone, la probabilità è generalmente indeterminabile.

Così nessuno può dire quale probabilità di morire nel corso dell'anno abbiano, singolarmente, Tizio, Cajo, Sempronio... che sono « nel mezzo del cammin di nostra vita »; ma per l'intera classe dei trentacinquenni diremo essere la probabilità del 10 ‰, se dalle osservazioni passate risultò che su 1000 uomini di quella età, *dieci* in media pagarono effettivamente il loro tributo alla morte nel corso dell'anno. Il passato si assume come norma approssimativa dell'avvenire.

Secondo le statistiche giudiziarie del periodo 1880-1899 in Italia, su 100 ricorsi, decisi in merito, contro sentenze di Corti d'Appello, se n'ebbero anno per anno almeno 33,5 e al più 39,6 — in media 36,6 — di cassazione della sentenza impugnata. Questa media può assumersi come espressione di probabilità solo per la massa indistinta dei ricorsi; chè, se si scendesse ad un'analisi preventiva dei singoli casi, se ne troverebbero di quelli in cui il motivo di cassazione è evidente o quasi, altri in cui è più o meno dubbio, ed altri ancora in cui è palese o pressochè palese la mancanza di una seria base. I singoli eventi non sono dunque possibili in egual grado e le rispettive probabilità potrebbero essere diversamente valutate da avvocati diversi.

Quindi concludesi: *La probabilità, che è generalmente indeterminabile per i casi singoli, diviene determinabile nella massa.* Ed è questo un terzo punto di antitesi fra i caratteri del fenomeno collettivo, come tale, e quelli dei casi singoli o dei gruppi scelti, di cui si compone.

Il calcolo di probabilità ha così estese ed interessanti applicazioni nell'àmbito dei fatti statistici che è prezzo dell'opera esporne gli elementi in apposito capitolo nella forma più piana possibile, dispensando il lettore, che se ne contenti, dal ricorrere a trattati speciali.

§ 7. L'idea di tipo, quella di costanza, continuità o periodicità e quella ancora di probabilità hanno una notevole influenza sulla nostra condotta. Gli uomini sono in piccola parte « originali », nella massima parte « imitatori » che conformano qual più, qual meno, i loro atti all'esempio dei primi; ond'è che ritroviamo il tipico persino nelle cose di più libera elezione, la foggia del vestire, l'ora del pranzo, il formato dei biglietti da visita, ecc. La periodicità riconosciuta di molti fenomeni influisce sulle disposizioni che prendiamo per fronteggiarli. Se il consumo di una merce si addimostra costante, vedrete

i produttori arrischiarsi ad investire capitale fisso in notevoli proporzioni nella loro industria; al contrario conviene ai produttori di merci di consumo assai variabile evitare le immobilizzazioni e lavorare con capitale circolante, facilmente disinvestibile. Gli istituti di assicurazione intendono a garantire la clientela loro dalle conseguenze economiche di avvenimenti probabili o anche di avvenimenti certi, dei quali però non si possa caso per caso prestabilire la data.

Ciò che è tipico in un sistema di varietà, risultando secondo ogni verosimiglianza da un concerto di cause in cui prevalgono quelle di carattere generale, riesce anche relativamente più stabile e probabile delle varietà singolarmente considerate. Nell'ordine biologico anzi, le variazioni delle specie animali o vegetali non si fissano se non con una rigorosa selezione artificiale o per via di qualche circostanza che assicuri il loro isolamento; se no, il tipo riappare, rinasce, dall'incontro delle specie variate in opposto senso, ma aventi (ecco la causa di carattere generale) comuni progenitori. In altri ordini di fatti, riducibili a combinazioni di elementi indipendenti, vi ha sempre una forma di combinazione che, in confronto delle altre, ha un maggior numero di casi favorevoli e quindi una maggior probabilità di verificarsi; è dessa che fornisce il tipo.

Ma la determinazione di ciò che è, relativamente parlando, tipico, costante, probabile, se riesce a mettere in evidenza l'aspetto formale del fenomeno collettivo, non esaurisce l'ordine di conoscenze che noi possiamo trarre dallo studio di esso. Tutto quanto è deviazione dal tipico o dal costante o dal più probabile in una massa di soggetti o di casi, ha pure la sua ragion d'essere, poichè è. Filosoficamente anzi l'importanza del principio delle variazioni è grandissima; e la Statistica potrebbe dirsi, come già fu detta l'Economia politica, un *ramo del calcolo delle variazioni*. I nostri modi di pensare, di volere, di agire corrispondono volta per volta a variazioni del mondo esteriore, che siano però abbastanza grandi da essere avvertite, che oltrepassino insomma la « soglia della coscienza »; se tutto intorno a noi fosse uniforme, stabile e certo, noi vivremmo la vita dei cristalli e non degli uomini. Se l'aria si raffredda, indossiamo abiti più pesanti o facciamo della novità l'argomento di conversazione spicciola o ci eleviamo a ricercare le cause della mutata temperatura: ecco variazioni di condotta determinate da variazioni esteriori. Se il raccolto riesce abbondante, rinforziamo le riserve, importiamo meno dall'estero o assistiamo al ribasso dei prezzi o prevediamo una mitigazione del cambio. E così via. Il motto « nulla si crea, nulla si distrugge » si applica bene a proposito di variazioni. Le stesse variazioni infinitesimali, che non arrivano direttamente al limitare della nostra coscienza, possono arrivarvi in maniera indiretta coi loro effetti lentamente accumulati e resi visibili all'occhio dell'osservatore. In breve, il problema delle « cause » che in Statistica già si affaccia, quand'è determinata la forma

generale del fenomeno, vien posto innanzi più decisamente dal fatto stesso delle sue particolari variazioni, che ne fanno supporre di antecedenti, concomitanti o susseguenti in altri ordini di cose.

§ 8. Ultimo intento è dunque quello di decomporre il sistema di cause o forze, di cui il fatto collettivo in esame sarebbe la risultante.

In meccanica si dice che una forza è completamente determinata, quando se ne conoscono i caratteri, cioè il *punto di applicazione*, la *direzione* e la *intensità*. La risultante di due forze concorrenti (le cui direzioni cioè s'incontrano in un medesimo punto, che può ritenersi il loro comune punto d'applicazione) è rappresentata, in grandezza e direzione, dalla diagonale del parallelogrammo costruito sulle rette, che figurano tali forze. Nell'ipotesi di un numero qualunque di forze applicate ad uno stesso punto, in direzioni e con intensità diverse, la risultante si ottiene applicando successivamente il principio anzidetto prima a due di queste forze, poi alla risultante ottenuta e a una terza forza e così di seguito fino all'ultima. In base allo stesso principio, data la risultante e una delle componenti, si giunge a determinare l'altra componente o la risultante dell'insieme delle altre componenti.

Queste nozioni fondamentali non si possono trasportare *sic et simpliciter* nell'àmbito dei fenomeni collettivi, a meno che non ci si voglia contentare di metafore o di analogie piuttosto vaghe. Gli è che in Statistica la parola « causa » o « forza » non ha e forse non può acquistare il senso ben determinato che le assegnano la meccanica, la fisica ed altre scienze. *Causa*, nel concetto di queste scienze, *è l'antecedente necessario o il sistema di antecedenti necessari per la produzione del fenomeno*. In quella vece, un concorso di antecedenti necessari *e di circostanze d'azione ignota o dubbia*, *non separabili da quelli*, non può più dirsi « causa », salvo in senso affatto generico, qual'è appunto accettato in Statistica. I fatti collettivi non comportano di solito che una determinazione di cause generiche; essi hanno, per meglio spiegarci, antecedenti che costituiscono essi medesimi dei fatti collettivi. Quando diciamo, ad esempio, che il tal fenomeno è funzione dell'« età » degli individui, che il tal altro dipende dalle « stagioni », che il tal altro ancora risente l'influenza delle « professioni », ecc., noi colle parole *età*, *stagione*, *professione* e simili enunciamo, non degli antecedenti semplici, ma dei complessi di circostanze, di cui alcune possono essere propriamente decisive per il verificarsi del fenomeno, altre invece indifferenti o contrastanti ed altre ancora di effetto non precisabile. L'età è l'espressione sintetica di un insieme di situazioni fisiologiche e psicologiche e di relazioni variabili col mondo esteriore; la stagione è pure la sintesi di fatti climatici, economici e così via. Per eliminare tutte le circostanze non necessarie bisognerebbe potere, come possono il fisico e il chimico, disporre gli sperimenti in condi-

zioni prestabilite nel modo più opportuno, isolare, combinare, variare a piacere certi elementi; ma l'osservatore statistico è forzatamente passivo di fronte agli obbietti del suo studio, manca della facoltà di riprodurre ad arte i fenomeni, di provocarli o di modificare a talento le condizioni in cui essi si verificheranno. Egli può solo specializzare l'indagine combinando in vario modo gli elementi della classificazione o scegliendo e circoscrivendo luoghi e tempi, per modo che una data circostanza abbia un massimo o un minimo d'importanza; ma anche lo specializzare e il circoscrivere incontrano presto, come vedremo, dei limiti pratici e teorici e, toccato il limite, ci lasciano daccapo in presenza di cause generiche, *sebbene meno generiche di prima.*

Questo carattere delle cause statistiche, per cui non figurano mai del tutto depurate da elementi estranei, accidentali o no, fa sì che sotto l'aspetto meccanico esse debbano considerarsi soltanto come forze imperfettamente determinabili, così nella loro intensità e direzione, come nei loro punti di applicazione. Per punto di applicazione vuolsi intendere il momento della serie in cui un nuovo fattore di variazioni viene ad introdursi nel concerto degli altri fattori, o anche quello in cui un fattore, già attivo, esce dal concerto di cui faceva parte. Conseguenza di ciò che abbiam detto è che le leggi statistiche e l'induzione statistica hanno un'impronta che le distingue nettamente dalle leggi naturali e dall'induzione sperimentale.

Se le cause, per il loro modo di operare, si riconoscono dagli effetti, è chiaro che la prima e più generale classificazione che di esse possiamo accogliere in Statistica deve trovarsi in accordo cogli aspetti formali dei fenomeni collettivi. Perciò, se una serie si segnala per la sua costanza o continuità, avremo presunzione legittima di *cause costanti e continue*, e nel caso particolare di serie periodiche, presunzione di *cause periodiche.* Le oscillazioni o vibrazioni minime della serie, che non impediscono all'occhio di cogliere l'andamento generale e regolare, suggeriscono la duplice categoria delle *cause variabili irregolari o accidentali* e delle *variabili regolari.* Toccherà poi all'analisi, fin dove la specializzazione dei dati lo consente, il dimostrare che in questo o quel caso concreto la continuità ha solo origine da una particolar forma di combinazione di fatti discontinui o che la uniformità d'andamento della serie nasconde effettivamente delle periodicità inverse e compensantisi presso i gruppi scelti, in cui la massa può frazionarsi.

L'ulteriore nomenclatura delle cause (*naturali*, *biologiche* e *sociali*, divise le prime in *telluriche*, *tellurico-cosmiche*, ecc., le seconde in *organiche*, *etniche*, ecc., le ultime in *economiche*, *politiche*, *intellettuali*, ecc., e tutte poi suddivise ancora in più categorie) non va, per importanza, al di là di un elementare schematismo, ai bisogni del quale sopperisce già abbondantemente la cultura generale d'ogni studioso.

§ 9. Per « legge » intendesi *la relazione costante che passa tra un fenomeno e la causa sua o una sua condizione generalissima, come il tempo e lo spazio.* Se la costanza è assoluta, si ha una legge « esatta »; se non è rigorosa o si verifica solo entro certi limiti, la legge si dice « approssimativa » o « empirica ».

Esatta o assoluta può, ad esempio, ritenersi la legge: *l'intensità della luce varia in ragione inversa del quadrato delle distanze.* Approssimativa invece la legge detta di Mariotte: *il volume di una data massa di gas è in ragione inversa della pressione che sostiene;* infatti questa legge non è più rigorosa quando i gas vengono sottoposti a una pressione vicina a quella che determina la loro liquefazione; inoltre è accertato che alcuni gas si comprimono un po' più, altri un po' meno di quel che la regola comporterebbe. È una legge approssimativa quella di Dulong e Petit, secondo la quale *i calorici specifici dei corpi semplici sono in ragione inversa dei pesi atomici.* A diverse latitudini o a diverse altezze sul livello del mare, il frumento richiederebbe tanto tempo per venire a maturanza, quanto occorre per formare, addizionando le temperature medie giornaliere, la somma di 2000 gradi di calore; in altri termini il tempo impiegato sarebbe in ragione inversa della temperatura media. Ecco di nuovo una legge approssimativa, anzi largamente approssimativa.

Le « leggi statistiche » non possono essere che approssimative, perchè, come si disse, le cause cui si riferiscono certi effetti sono più o meno generiche e cioè contengono in dosi variabili, insieme agli antecedenti necessari, dei fattori di azione ignota o dubbia non separabili da quelli. Per quanto si specializzino le masse in gruppi scelti aventi a comune parecchie caratteristiche fondamentali, noi ci troviamo, prima come poi, in presenza di quantità essenzialmente variabili. La determinazione di ciò che è tipico, costante o più probabile, riposa su processi dai quali non ci è dato di bandire l'elemento soggettivo e cioè una certa misura d'arbitrio. Pertanto le leggi statistiche debbono intendersi nel senso discreto di leggi non estendibili al passato non osservato o all'avvenire maturante sempre fatti nuovi — salvo in via di semplice probabilità — oppure come relazioni empiriche valide per la massa dei casi e non per i casi singoli. A rigore si dovrebbero dire *interferenze complesse di leggi esatte, presentantisi con tali caratteri di uniformità da* PARERE *esse stesse delle leggi semplici.*

Della parola « legge » si fa un abuso nelle scienze sociali, giustamente riprovato dal Rümelin (1). Gli è che si confonde il concetto di legge, da un lato con quello di principio, di teorema, ecc., dall'altro lato colle semplici nozioni empiriche delle uniformità di natura. In

(1) G. Rümelin, *Sul concetto di una legge sociale* (*Annali di Statistica*, serie 2ª, vol. XXIII). Traduzione di V. Cusumano.

generale le relazioni che si discoprono per via deduttiva o si ammettono per astrazione sono, secondo i casi, principii, teoremi, ipotesi, ecc.; quelle alle quali si giunge per via induttiva sono leggi. Se ci si può arrivare per l'una o per l'altra via indifferentemente, si userà ancora la prima nomenclatura, avendo le verità deduttive un carattere più assoluto; l'uso promiscuo dei termini è ammissibile solo in pochi casi. Che la somma degli angoli di un triangolo sia costantemente eguale a due retti è un teorema o una proposizione deducibile a tutto rigore da certe proposizioni fondamentali; non una legge. Sarebbe una legge, se per giungere ad essa non si fosse conosciuta altra via che quella di misurare un gran numero di triangoli diversi l'uno dall'altro. La relazione causale si esprimerebbe allora così: « se lo spazio intercettato da due rette concorrenti vien chiuso da una terza (antecedente causale) comunque condotta, la somma degli angoli (conseguente) è *per osservazione* una quantità costante ed eguale a due angoli retti ».

Correttamente devesi pur dire « principio della popolazione secondo Malthus » e non legge della popolazione; « principio della rendita ricardiana » e non legge della rendita ricardiana; « principio delle probabilità » e non leggi di probabilità e molto meno *leggi del caso!*

Importa pure distinguere tra « leggi » e « nozioni empiriche concernenti l'uniformità o regolarità dei fenomeni ». Il sapere che nell'anno tale e tale altro si ebbero eclissi di sole o di luna, è una *nozione semplice;* sapere che ad ogni ciclo di 223 lunazioni (ossia 18 anni, 11 giorni e 7-8 ore) si ha una ripetizione assai approssimata della serie di eclissi precedente, vuol dire avere la *nozione empirica di una uniformità di natura;* determinare infine gli antecedenti causali di questa periodicità, ricollegarla alle leggi della meccanica celeste, è affermare una *legge.*

Del pari sapere che nelle famiglie possono darsi figliuolanze composte unicamente di maschi o unicamente di femmine oppure in modo vario assortite, è una nozione empirica semplice. Avere osservato che su un gran numero di famiglie la somma delle nascite maschili sta a quella delle femminili nel rapporto abbastanza costante di 106 a 100, significa possedere la nozione empirica di una uniformità di natura; sapere infine (se per ipotesi ciò si scoprisse davvero) che, date certe combinazioni d'età degli sposi, un certo regime alimentare è decisivo per la nascita di un maschio, sarebbe conoscere una legge.

L'osservazione comune non frutta di solito che nozioni semplici; la metodica (osservazione sperimentale, statistica e storica) estende le nostre cognizioni sulle « uniformità » della natura; l'induzione lavora a convertire queste nozioni empiriche di uniformità in leggi vere e proprie.

Statistici e non statistici sogliono parlare di una legge « dei grandi numeri ». Che cosa vogliono significare? Questo soltanto, che accrescendosi la massa delle osservazioni, si accresce la probabilità che

le cause costanti e continue prendano il sopravvento sulle accidentali e discontinue, le cause generali sulle particolari. Ma qui, meglio che una legge, è in questione un principio o un corollario dei principii delle probabilità, siccome vedremo a tempo debito. Non diversamente devesi dire della legge « dei piccoli numeri », elegante novità della statistica matematica, dovuta al Bortkewitsch (1), della quale pure parleremo.

§ **10.** Altri concetti generali, che spesso ricorrono in Statistica, sono quelli di rapporto, correlazione, interdipendenza, massimo e minimo di una funzione, limite.

Rapporti statistici non si possono istituire indifferentemente fra termini presi ad arbitrio. Dividere, ad es., il numero delle stelle visibili ad occhio nudo per quello degli abitanti di un paese, non avrebbe senso. Nemmeno l'avrebbe il dividere i quintali di grano importati per il numero degli infortuni sul lavoro. Occorre invece, a base del confronto aritmetico, una relazione logica come di parte al tutto, di effetto a causa, di contenuto a contenente, di situazione statica a dinamica, di manifestazione eccezionale a manifestazione normale, e così via.

L'idea di rapporto che involge evidentemente quelle più generali di somiglianza e di opposizione fra due termini, non deve essere confusa coll'idea, per altri rispetti affine, di *correlazione.*

« Rapporto » è comparazione di grandezza fra due quantità concrete, considerate in sè stesse, in quanto opponibili l'una all'altra, come la parte al tutto, l'effetto alla causa, ecc., ma sempre astrazione fatta dalla loro capacità di variare l'una in funzione dell'altra. Non c'è bisogno che i termini posti a confronto siano rappresentati da due *serie* di numeri; basta che ciascuno lo sia da *un* numero. « Correlazione » invece implica l'idea di « variazioni concomitanti » e quindi di serie numeriche. Può darsi anzi un rapporto statistico tra due fatti senza correlazione assegnabile tra loro, come può darsi correlazione senza rapporto. Esempio: in Italia il debito pubblico dello Stato è di 400 lire circa per abitante. Ecco un rapporto. Ma correlazione non c'è o può non esserci; vale a dire, non è dimostrato che ad ogni variazione in più o in meno nel numero degli abitanti corrisponda con sufficiente costanza una variazione nello stesso senso (o nel senso contrario) dell'ammontare del debito pubblico. Facciamo ora un caso di correlazione senza rapporto. Non si dà rapporto fra il numero dei matrimoni, come tale, e il prezzo del grano, come tale; nè avrebbe significato alcuno il dire: in Italia nell'anno 1901 i matrimoni furono 235.000; il prezzo del grano fu di lire 20 l'ettolitro; *ergo*, 11.750 coppie

(1) Dott. L. von Bortkewitsch, *Das Gesetz der kleinen Zahlen*, Leipzig, B. G. Teubner, 1898.

di sposi per ogni lira di prezzo della derrata! Ma correlazione tra i due fatti può esserci e c'è effettivamente. Se il grano rincara del 20, del 30, del 50 % sul prezzo ordinario, non ci meraviglieremmo di sentire che la nuzialità, in conseguenza appunto del rincaro dei viveri, diminuisce di un tant'altro per cento.

Interdipendenza esiste fra due fenomeni, allorquando quello dei due che si considera come effetto dell'altro reagisce sulla causa sua modificandola. Così, l'aumento della popolazione può avere per effetto una maggiore agglomerazione delle famiglie nelle case e questa agglomerazione favorisce da un lato unioni precoci o irregolari, donde deriva ancora aumento di popolazione, dall'altro lato epidemie che la diminuiscono. Il crescere della domanda di una merce per parte di alcuni strati di consumatori provoca un rialzo del prezzo; ma il prezzo rialzato arresta la domanda di costoro e la diminuisce negli altri strati. Tale reciprocità d'azione e di reazione seguita fino ad un punto che possiam dire di *equilibrio stabile*.

Il concetto di *funzione* in Statistica non differisce da quello accolto nelle matematiche. Quando i valori successivi di una quantità variabile dipendono, secondo una certa regola, da quelli che assume un'altra variabile, la prima si dice una « funzione » della seconda. L'area di un circolo si dirà dunque funzione del suo raggio; crescendo il raggio come i numeri 1, 2, 3..., l'area cresce corrispondentemente come 1, 4, 9..., cioè come i quadrati dei raggi. La durata delle oscillazioni del pendolo è funzione della lunghezza di questo. Se la lunghezza diventasse 4, 9, 16... volte maggiore di quella che si è presa come unità di confronto, la durata delle oscillazioni crescerebbe come la radice quadrata di cotesti numeri, cioè sarebbe 2, 3, 4... volte maggiore. Similmente in antropometria diciamo che la frequenza del polso è una funzione della statura; in economia, che il consumo di una merce è funzione del prezzo, ecc. I casi di correlazione e di interdipendenza rientrano nel concetto più ampio di funzione.

Una quantità può essere funzione di più variabili indipendenti. Per esempio, il volume di un cilindro è funzione della sua altezza e del raggio della sua base; la fecondità è funzione dell'età dell'uomo e della donna, ecc.

Se, facendo crescere la variabile, la funzione prende un valore reale, che sorpassa i valori che immediatamente lo precedono o gli susseguono, quel valore dicesi un *massimo* della funzione. Dicesi invece *minimo*, quando il valore trovato è inferiore a quelli che immediatamente lo precedono e lo seguono. Così, per un venditore monopolista il prezzo più conveniente è quello che moltiplicato per la quantità esitabile sul mercato gli assicura un massimo prodotto netto. Un prezzo più elevato al pari di uno meno elevato gli ridurrebbe il prodotto netto, con suo scapito. Lo sviluppo della forza renale o di sollevamento, come funzione dell'età, raggiunge il suo massimo a 30 anni;

indi declina. Naturalmente poi, tra fenomeni connessi per qualche rapporto causale havvi corrispondenza dei rispettivi momenti di massimo o di minimo. Ad esempio, nell'età in cui l'operaio arriva al culmine di sviluppo della forza fisica, arriva pure al *maximum* di produttività economica e, per riflesso, di salario, almeno nelle professioni manuali più semplici — ecco un caso di corrispondenza.

Nei fenomeni collettivi rappresentati in serie, le oscillazioni secondarie prodotte da cause perturbatrici d'ordine accidentale formano dei massimi e minimi apparenti e numerosi; i massimi e minimi reali della funzione non si possono individuare che in maniera approssimativa sulle curve interpolatrici, che figurano o si suppongono figurare l'andamento generale del fenomeno.

L'idea di *limite* è famigliare alle matematiche e alle scienze fisiche e va diventandolo nelle discipline sociali. Quando i valori successivi di una quantità variabile si avvicinano indefinitamente a una quantità fissa e determinata, in maniera da differirne tanto poco quanto si desidera, questa quantità fissa si chiama il « limite » dei valori della variabile. Il numero 2 è, per es., il limite cui tende la somma $1 + 1/2 + 1/4 + 1/8 + 1/16$, ecc. La superficie del circolo è il limite verso il quale tende l'area del poligono regolare inscritto, man mano che si accresce il numero dei lati. In Demografia si parla di un limite segnato dalle sussistenze all'aumentare della popolazione; in Economia di terre-limiti, cioè le terre meno fertili o più lontane dal mercato, la cui coltivazione è ancor necessaria e che dànno solo l'ordinario profitto, mentre le altre terre, più fertili o più vicine al mercato, dànno ad un tempo profitto e rendita ricardiana. Nei rapporti statistici (come la proporzione delle nascite maschili alle femminili, la proporzione delle sentenze appellate sul totale delle sentenze pronunziate da una magistratura di prima istanza, e simili) che si assumono come espressione di una probabilità costante, si parla pure, ma in un senso speciale, di *limiti probabili* d'oscillazione dei rapporti stessi. Questa nozione, per la sua stessa specialità, non può essere chiarita ora, in sede di nozioni generali; ci conviene riservarla al capitolo che avrà per argomento la teoria delle probabilità e le sue applicazioni ai fatti collettivi.

Passiamo a dire della Statistica come ramo della Logica.

CAPO SECONDO

La Statistica come ramo della Logica.

Sommario: § 1. Forme di osservazione. — § 2. Forme di ragionamento. — § 2. La Statistica come svolgimento particolare del metodo induttivo. — § 4. Separazione della Statistica metodologica dalla *Demografia*, scienza quantitativa della popolazione, e dalle *Descrizioni statistiche di fenomeni sociali*.

§ **1.** La definizione da noi data della Statistica ce la indica come una forma particolare di osservazione e di induzione appropriata allo studio dei fenomeni collettivi. Dobbiamo dunque rispondere alla domanda: quale posto tiene l'osservazione statistica tra le forme di osservazione? e quale l'induzione statistica tra le forme di ragionamento?

L'osservazione, se ha per obbietto le cose o i fatti del mondo esterno, suppone l'esercizio dei sensi; se riguarda i fenomeni interiori del soggetto pensante suppone semplice attività di coscienza. L'esercizio dei sensi, quando non è assistito da speciali mezzi che ne accrescano la portata e la precisione o non si conforma a regole costanti per la scelta degli oggetti e delle condizioni opportune d'esame, dà luogo all'*osservazione comune*. In caso diverso si ha l'*osservazione metodica*. Quella è limitata al breve raggio d'esperienza dell'individuo, è discontinua, variabile come le condizioni soggettive di esso, come gli stati d'animo; non discende sotto quello che in psicologia dicesi *minimum sensibile* o la soglia della coscienza (1); muore spesso coll'individuo o gli sopravvive più o meno alterata dalla tradizione orale. La metodica invece riesce senza confronto più autorevole o per l'uso di strumenti a correzione ed aiuto degli imperfetti sensi naturali o per la colleganza ed il controllo dei risultati ottenuti da diversi osservatori o per la continuità nel tempo assicurata da mezzi permanenti (scrittura, stampa) e da organi d'indagine costituiti a forma di pubblica funzione, o per la disciplina di regole logiche nella scelta dei momenti, luoghi e modi di accertamento dei fatti. Ora la Statistica appartiene senza alcun dubbio alle forme di osservazione metodica.

---

(1) Le piccolissime variazioni dei fenomeni sfuggono ai nostri sensi; noi non avvertiamo la differenza di pochi grammi aggiunti o tolti a nostra insaputa a un chilogramma sostenuto colla nostra mano; non la differenza di qualche decimo di grado nella temperatura dell'ambiente, ecc. Così, senza la enumerazione precisa dei casi, ci sfuggirebbero le piccole variazioni di mortalità da un mese all'altro, dall'una all'altra settimana; bisogna che il fenomeno prenda le proporzioni di una epidemia per arrivare a notizia del comune osservatore.

Quanto all'oggetto distinguesi l'osservazione secondo che si porta su *fenomeni singoli* o su *fenomeni collettivi considerati come un tutto unico, come unità in grande*, ovvero ancora su *fenomeni collettivi considerati semplicemente come pluralità variabili di casi.* Se, per esempio, dell'individuo tale o tal altro faccio oggetto di speciale esame antropometrico, fisiologico, psicologico, ecc., avrò la prima delle tre forme ora accennate di osservazione; se invece studio la popolazione di un paese nelle sue vicende storiche, nello spirito delle sue istituzioni, credenze, ecc., sempre trattandola come unità in grande e prescindendo dagli aspetti numerici che può assumere il fenomeno, avrò la seconda forma; infine se di questa stessa popolazione enumero, classifico, graduo i componenti secondo il sesso, l'età, lo stato civile, il luogo di nascita, ecc., confronto gruppi e classi, situazioni attuali e passate e così via, avrò la terza forma di osservazione, precisamente quella che dicesi statistica.

V'è poi, dal punto di vista del soggetto che l'intraprende, l'osservazione *individuale* e la *associata* o *cooperativa.* L'individuale può avere grande valore, se chi la compie ha qualità eminenti di osservatore, massime nelle discipline sperimentali. Ma quante volte si richiedano simultaneità e coordinamento di osservazioni in parecchi luoghi o mezzi eccedenti le ordinarie facoltà di un privato o quando la breve durata di un fenomeno complesso non permetta ad un solo osservatore l'esame specializzato, come si conviene, dei diversi caratteri, la divisione del lavoro s'impone. Una eclissi di sole, osservata in parecchi punti del cono d'ombra da differenti astronomi, dura in certo modo più tempo che non duri per uno solo tra essi; la probabilità che il cielo sia coperto e faccia fallire l'osservazione a tutti, è molto minore che per uno solo. L'osservazione statistica è raramente individuale; se lo è non può aggirarsi che su fatti di poca estensione, ma in compenso suscettibili d'indagini intensificate, specializzate; di solito la ritroviamo stabilita sulla base della cooperazione, anzi bene spesso organizzata a mo' di pubblica inchiesta. Solo in quest'ultima maniera si rendono possibili le colossali operazioni dei censimenti, la regolarità dei rilievi circa il movimento della popolazione (nascite, morti, matrimoni, migrazioni) e le notizie numeriche che illustrano molt'altri fatti collettivi, naturali o sociali.

Da ultimo suolsi distinguere l'osservazione in *semplice* e *sperimentale*, secondo che si esercita attorno al fenomeno senza padroneggiarlo o invece riesce a riprodurlo artificialmente; in *attuale* e *storica*, secondo che guarda all'oggetto nelle manifestazioni del momento o invece ne considera le fasi attraversate, la evoluzione, come dicesi, nel tempo. Il fisico e il chimico fanno quasi sempre dell'osservazione sperimentale, essendo in loro potere di variare le circostanze in cui si produce un fenomeno, ad es., col predisporre una data pressione, temperatura o cangiare le proporzioni di una soluzione, stabilire contatti, ecc. La

osservazione storica si può dare in ogni campo del sapere, purchè i fenomeni del passato abbiano lasciato traccie sufficienti di sè nel presente. Come le sedi di antichi popoli si riconoscono dai nomi caratteristici di certe località e i tipi etnici dalle misure sui cranii, ecc., così si arriva a stabilire che il clima, poniamo, della Sicilia era diverso all'epoca romana da quel che è oggi, raffrontando i passi degli scrittori che accennano a ricchezza d'acque correnti o a specie animali e vegetali, per le quali è condizione di vita un clima umido anzichè secco. Qui però, meglio che di osservazione, dovrebbesi correttamente parlare di induzione storica. Quanto alla Statistica, ell'è senza dubbio una forma di osservazione semplice e non sperimentale, che ripetendosi nel tempo prende le sembianze di una « storia in cifre », ma che può anche arrestarsi alla manifestazione attuale dei fenomeni collettivi.

Alcuni scrittori escludono l'osservazione dal dominio della Logica, per due motivi principali: anzitutto, perchè essa in ultima analisi si risolve in pure intuizioni e conoscenze immediate; in secondo luogo perchè le sue regole non si possono generalizzare e variano dall'una all'altra scienza, ciascuna scienza essendo la sola competente a dettarle per sè. La Logica pertanto si restringerebbe alla *definizione*, *induzione* e *deduzione* (1). Noi riteniamo invece che il primo motivo valga solo per l'osservazione comune, non per la metodica, la quale implica di necessità delle operazioni mentali antecedenti e indipendenti da quelle che serviranno poi a generalizzare i casi osservati per induzione o a ricavare per deduzione nuove verità particolari. Queste operazioni mentali sono in sostanza dei giudizi di scelta tra i diversi mezzi o luoghi o momenti che si presentano per osservare un fenomeno. Quand'anche noi non volessimo affermare nulla mai al di là dei limiti della nostra effettiva esperienza passata — l'affermare al di là di questi limiti è caratteristica dell'induzione — l'osservazione metodica rimarrebbe il fondamento logico delle scienze descrittive. Anzi, in tale ipotesi, non ci sarebbero che scienze descrittive; ma i loro metodi formerebbero, per quanto povera, materia di una Logica. Riguardo poi al secondo motivo, si risponde che se le regole della osservazione non si sono ancor potute generalizzare, ciò dipende forse dalla scarsa attenzione che a questo obbietto hanno prestato gli autori (2); mentre non è improbabile che presto o tardi si discoprano tanti elementi comuni nelle regole delle diverse discipline positive da formare un sistema, la cui sede naturale non potrà essere che la Logica. Frattanto, la Statistica non offre già un saggio di sistemazione dei

---

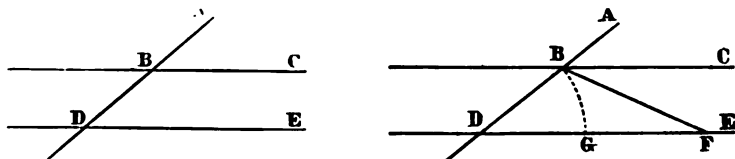
(1) Veggasi A. Bain, *Logique déductive et inductive*, vol. I, pag. 53, 55, 59, Paris, Germer Baillière, 1881.

(2) Il Bain, op. cit., pag. 53, dice: « De tous les procédés essentiels de la logique, l'observation est celui qui est généralement le moins étudié ».

procedimenti d'osservazione dei fenomeni collettivi, a qualunque ordine di conoscenze appartengano?

Noi ammettiamo quindi che la Statistica, come forma di osservazione metodica di certi fatti, entri già a questo titolo nella sfera della Logica. Se è vero che ogni disciplina speciale è la sola competente a dettare le proprie particolari norme di osservazione, ci sembra inevitabile ammettere che anche la Logica, come scienza generale delle forme della prova e dell'evidenza, sia competente a scrivere dell'osservazione le regole generali.

§ 2. Osservare non è ancora « indurre », nè molto meno « dedurre ». *Indurre* significa affermare che ciò che fu constatato vero per certi casi osservati, è vero per tutti i casi della stessa categoria. Ad esempio: constatato che per *alcuni* corpi, qualunque sia l'intensità della sorgente di calore, dal momento in cui la fusione incomincia a quello in cui si compie la temperatura cessa di innalzarsi, si può concludere « induttivamente » che ciò sarà vero per *tutti* i corpi in corso di fusione.

Dedurre significa affermare che ciò che è vero o si ammette per vero come principio generale, è vero o deve ammettersi per vero nei casi speciali qualificati da condizioni, che non alterino quella essenziale

per cui sussiste il principio. Così, stabilito per antecedente dimostrazione che una retta, la quale taglia due altre, fra loro parallele, forma gli angoli corrispondenti eguali (l'angolo ABC=ADE), questa eguaglianza sussisterà pure nel caso particolare in cui l'angolo ADE, invece di essere aperto, sia chiuso da una retta come BF, oppure da una curva come BG, essendo indifferente per qualificare la grandezza dell'angolo che lo spazio intercettato sia chiuso piuttosto che aperto.

Bisogna distinguere l'induzione dalle *semplici generalizzazioni*. Se dopo avere constatato *per ogni singolo pianeta* che esso splende di luce solare riflessa, concludiamo che *tutti* i pianeti conosciuti splendono di luce solare riflessa, noi non facciamo che riassumere i singoli risultati in uno ed esprimerli, anzichè in tante distinte proposizioni particolari, in una proposizione generale. Ma la nostra non sarebbe una induzione, cioè una estensione di ciò che si è trovato vero per *alcuni casi osservati* ai *casi non osservati* della stessa categoria. Perchè si potesse parlare di induzione, bisognerebbe che il fatto della luce riflessa fosse accertato solo per *alcuni* pianeti e affermato poi senz'altro anche per i rimanenti.

La fecondità del ragionamento induttivo dipende dalla nostra attitudine a riconoscere che i casi non osservati, cui si vuole estendere la regola dei casi osservati, sono veramente dello stesso ordine di questi, e cioè, per quanto qualificati dalle più diverse apparenze, non contengono tuttavia condizioni per cui debba venir meno la regola o il principio. Ne segue che le generalizzazioni ammesse per via induttiva costituiscono verità assolute, solo quando affermano che ciò che si è verificato in un caso si verificherà per lo stesso caso ripetuto nei precisi limiti dell'esperienza già fatta. Se sconfiniamo idealmente da questi limiti o se estendiamo la regola da caso a caso della stessa varietà, da varietà a varietà della stessa specie, da specie a specie dello stesso genere, le verità indotte sono approssimative. L'approssimazione è spesso grande tanto da confondersi colla certezza pratica nelle scienze fisiche, dove ogni nuova esperienza viene a confermare l'uniformità o tipicità di certe manifestazioni della natura; ma in altri ordini di fenomeni l'approssimazione è più o meno lata. Diceva bene il Laplace, che quasi tutte le nostre conoscenze non sono che probabili.

L'osservazione rimane, in ultima analisi, il criterio migliore della attendibilità delle nostre generalizzazioni induttive. È dessa che decide se le condizioni speciali dei nuovi casi sono o non sono indifferenti per il verificarsi del fenomeno secondo la regola già indotta. I primi sperimentatori, i quali notarono che un raggio di luce, incontrando sotto un qualsiasi angolo una superficie levigata, si riflette facendo un angolo eguale a quello d'incidenza e che il raggio così riflesso, cadendo alla sua volta su una seconda lastra, si riflette di nuovo secondo la stessa legge, poterono credere alla validità assoluta della regola; tuttavia nel 1810 il Malus trovò che il raggio riflesso sotto un certo angolo *si polarizza*, ossia diviene inetto a riflettersi una seconda volta in date direzioni. Lo sperimento così ha dimostrato che certe condizioni, le quali si sarebbero potute ritenere indifferenti, invece non lo erano.

La fecondità del ragionamento deduttivo dipende invece dalla nostra attitudine a combinare idealmente parecchi principii generali e a mostrare come coesistono nel caso particolare in esame. Tutta la geometria, ad esempio, non è che un processo logico di successive combinazioni di verità intuitive o volta per volta dimostrate, le quali concorrono appunto nella fattispecie. Così dicasi della meccanica razionale, di gran parte dell'ottica e della stessa economia politica. A differenza dei processi induttivi, che solo presuppongono la esperienza formata della regolarità delle cose in natura, i deduttivi presuppongono una particolare attività d'immaginazione o invenzione.

Non possiamo ristarci dal dire il nostro avviso su una elegante questione sollevata a proposito dell'unità delle operazioni logiche. La deduzione si ridurrebbe, secondo il Mill, il Bain ed altri propugnatori di questa unità, ad una *induzione dissimulata*, cioè ad una semplice

estensione di ciò che è stato trovato vero per certi casi osservati ad altri casi non osservati della stessa classe. Il ben noto sillogismo: « tutti gli uomini sono mortali; Tizio è uomo; dunque Tizio è mortale », preso come tipo di ragionamento deduttivo, si risolve in realtà in una proposizione induttiva: « tutti gli uomini *del passato* risultarono mortali; questa qualità di « mortali » *si estende* (ecco l'induzione) anche ai casi non osservati, cioè agli uomini attualmente vivi e a quelli che vivranno, non escluso Tizio, che è vivo o che verrà al mondo ».

Ciò è vero, ma è vero soltanto perchè il sillogismo ricordato è tutt'altro che un tipo ideale di ragionamento deduttivo; quindi siamo ben lontani da una scoperta, che alcuni magnificano come una rivoluzione portata nella Logica. Rigorosamente parlando, il ragionamento deduttivo si ha quando il punto di partenza non è propriamente un fatto d'osservazione, ma un'*ipotesi* o un'*astrazione* o una *situazione-limite* in una data direzione delle nostre conoscenze, qualche cosa insomma che implichi quella particolare attività inventiva cui si accennava poc'anzi. Gli è da questa meravigliosa facoltà dello spirito che son venuti fuori i concetti dell'infinitesimo nel calcolo superiore, del punto inesteso in geometria, dell'atomo in chimica, dell'etere in fisica, del corpo perfettamente rigido in meccanica, dell'*homo oeconomicus* in economia politica, ecc. È evidente che qui si prescinde o anche si trascende, quanto al principio, dalla possibilità presente o futura dell'osservazione. Combinando poi la premessa maggiore con premesse minori o verità particolari, anche di carattere induttivo, si arriva ad una costruzione o ricostruzione puramente ideale di fenomeni, che ha un valore logico tanto più grande, quanto più grande è l'accordo colla realtà. Appunto perchè si tratta di costruzioni o ricostruzioni ideali, noi preferiamo chiamare *principii*, *teoremi*, ecc., le proposizioni che sono il punto di partenza del ragionamento deduttivo o che da esso discendono, mentre riserbiamo il nome di *leggi* alle generalizzazioni induttive.

Tra i molti classici esempi di vere e proprie deduzioni non v'è che l'incomodo della scelta. Prendiamo a caso. Il Thünen, per istudiare l'influenza del costo del trasporto sulla distribuzione topografica delle colture agrarie, immagina (si noti bene: immagina) uno Stato isolato di forma circolare, di produttività uniforme, con un sol mercato di consumo al centro e, possiamo aggiungere, abitato da « edonisti perfetti ». Questo non è certo un principio generale indotto, ma una ipotesi, un'astrazione. Lo stesso concetto di *homo oeconomicus* o edonista perfetto è l'idealizzazione di una qualità, che si trova in differentissimi gradi negli uomini reali e che vien mentalmente staccata da ogni altra qualità o carattere, nonostante che gli uomini in carne ed ossa ci si presentino come combinazioni indecomponibili di caratteri, fra cui esistono correlazioni e interdipendenze in gran parte

oscure. Ma, per tornare al Thünen, ricorderemo anche le premesse minori della teoria, cioè che i varii prodotti agricoli hanno un diverso peso e volume a parità di valore e un grado diverso di conservabilità; che il costo di trasporto cresce col crescere della distanza e, a parità di distanza, cresce col crescere del peso della merce, ecc. Combinando convenientemente premesse maggiori e minori, il Thünen conclude che un *maximum* di economia si realizza, quando i prodotti più pesanti, a parità di valore, o i più difficili a conservarsi, sono coltivati in prossimità del mercato di consumo e quando per prodotti della stessa specie, suscettibili di coltura più o meno intensiva, le colture si dispongono in anelli circolari tanto più eccentrici, quanto più estensivi. L'accordo tra il fenomeno così idealmente ricostruito e la realtà è a questo punto assai imperfetto; ma nulla impedisce che lo si renda perfetto, complicando via via le premesse col supporre, ad esempio, due o più mercati di consumo in luogo di uno solo, zone di maggior fertilità per certi prodotti e zone più adatte a certi altri, e così di seguito.

Sopra abbiamo accennato a « situazioni-limite » in varie direzioni delle nostre conoscenze, le quali possono essere assunte come premesse di veri e proprii ragionamenti deduttivi. Ciò che si vuol significare con quel termine, meglio che da una definizione, s'intenderà da un caso, tra i molti, che ci presta la geometria applicata alle scienze fisiche. È una elegante questione di ottica. Le lenti convergenti ordinarie (non acromatiche) si possono paragonare a serie di prismi, a faccie infinitamente piccole, riuniti colle loro basi. Questa, diciamolo subito, non è una verità induttiva, un risultato di osservazioni che sia stato esteso da casi accertati a casi non accertati, ma è piuttosto un modo di rappresentarci una situazione-limite, che trascende la possibilità dell'osservazione. Per quanto si impiccioliscano e si moltiplichino, fisicamente parlando, i prismi così disposti in serie, tra il loro insieme e la struttura di un corpo biconvesso a curvatura perfetta ci sarà sempre un salto, un *hiatus*, che solo la nostra immaginativa sa colmare. La identificazione, che noi facciamo per questa facoltà dello spirito, può allora porsi a base di un ragionamento deduttivo. Sapendosi dunque che il prisma rifrange e decompone la luce, concluderemo che anche le lenti convergenti ordinarie oltre al rifrangere la luce, la decompongono. Tale la spiegazione del fenomeno conosciuto sotto il nome di « aberrazione di rifrangibilità » per il quale vediamo gli oggetti con contorni iridescenti e che viene eliminato con processi speciali nelle lenti acromatiche.

§ **3.** Dal seguito dell'opera apparirà più chiaro ai lettori, se la Statistica possa considerarsi come un metodo *sui generis* distinto dall'induttivo e dal deduttivo o invece si riduca ad uno svolgimento particolare di uno dei due. Qui ci limiteremo a dire che le proposi-

zioni sue sono vere in generale, come media dei casi osservati, ma possono non essere vere nei singoli casi particolari. Dal fatto, per esempio, che le cambiali, che si creano in Italia, hanno un valor medio di 700 lire, non segue necessariamente che la cambiale emessa da Tizio o da Caio sia dell'ammontare giusto di 700 lire. Però l'affermazione correrebbe in questa forma: « Il medio importo degli effetti cambiari in Italia è di 700 lire; dunque *prendendo a caso alcune migliaia* di tali effetti, troveremo differire assai poco il loro medio ammontare dalla cifra di 700 lire ». In altri termini, da un enunciato vero in generale, come media di moltissime e differenti situazioni, si può discendere ad una applicazione d'una generalità più limitata, se nella fattispecie concorrono queste due condizioni: 1° che le cambiali, a cui si restringe l'osservazione, siano prese *a caso;* 2° che lo siano in numero abbastanza considerevole da rendere improbabile un risultato divergente troppo dalla media generale.

Si può invertire la deduzione e farla diventare induzione. Così, se presi a caso 5 o 10 mila effetti, trovo che il loro medio importo è di 700 lire, concludo che questa cifra rappresenta con una certa probabilità il valor medio di tutte le cambiali create in paese.

Delle due condizioni sopra enunciate, la seconda non è, come vedremo, che un corollario del calcolo di probabilità, mentre la prima deriva già da ciò che sappiamo intorno ai caratteri dei fenomeni di massa in opposizione a quelli dei casi singoli o di gruppi scelti. Si ricordi: « gli elementi tipici per gruppi scelti cessano di essere tipici nella massa, nella quale si comportano piuttosto come casi di deviazione dalla norma generale ». E viceversa: « L'elemento, che è tipico per la massa, può figurare come caso di deviazione nei gruppi scelti di questa ». Dunque nell'esempio di cui sopra, la condizione che le cambiali siano prese a caso s'identifica con quest'altra, che esse *non* costituiscano gruppi scelti, perocchè ciascun gruppo scelto si distingue per un proprio valore o ammontare tipico, più o meno divergente dalla media generale. La cambiale dell'operaio o del contadino sarà di minor conto di quella del piccolo industriale o agricoltore; la cambiale del piccolo industriale o agricoltore non arriva all'altezza di quella del grande, meno ancora a quella del banchiere, ecc. Gli effetti dei non commercianti si distinguono per una propria media diversa da quella degli effetti dei commercianti individui e più ancora delle società commerciali. In breve, tanti gruppi scelti, tanti valori tipici.

Senza più anticipare su quanto si dirà in proposito a tempo debito, concluderemo per ora che la Statistica, già al suo primo aspetto, ci si presenta non come un metodo diverso dall'induttivo e dal deduttivo, ma come uno svolgimento particolare dell'uno e dell'altro, anzi soprattutto del primo. La qual cosa abbiamo voluto che risaltasse nella definizione, ove appunto si è qualificata la Statistica come una *forma di induzione*, senza aggiungere: *e di deduzione*, appropriata

allo studio dei fenomeni collettivi. Infatti non è che il ragionamento deduttivo non abbia una certa parte in questo studio; ma esso ha un ufficio ausiliare soltanto o sta a testimoniare l'intervento di concetti tratti da altre discipline, in ispecie dalle matematiche. Invece la singolarità della Statistica, come ramo della logica dei metodi, appare tutta nel processo induttivo e nel fondamento di questo, che è l'osservazione.

Anche le parole « *appropriata allo studio dei fenomeni collettivi* » avevano la loro ragion d'essere nella definizione. Non esiste una Logica metafisica, una legge dell'Idea in sè considerata, al di fuori insomma delle sue determinazioni empiriche, come pretendono gli Hegeliani; è la condotta del pensiero che deve accomodarsi ai diversi ordini di fatti, non questi a quella. La Logica è la scienza teorica e pratica della prova e dell'evidenza; essa insegna a distinguere l'evidenza completa dall'evidenza sufficiente e dall'insufficiente o, in altri termini, la certezza dalla probabilità e dall'assurdità nelle forme del nostro ragionare. L'indole dei fenomeni collettivi ammette solo una soluzione approssimata o probabile del problema delle cause; però lo strumento del numero conferisce a questa probabilità una determinatezza, che manca invece ad altri ordini delle nostre conoscenze.

§ **4.** Il fatto che le migliori scoperte acquisite col metodo statistico si ebbero nella categoria dei fenomeni della popolazione, ha confermato molti scrittori nell'idea che la Statistica sia ad un tempo metodo d'investigazione e scienza autonoma avente un proprio oggetto di studio: gli *aggregati sociali*, un punto di vista suo particolare per la trattazione: quello del *numero*, e una meta ben precisata: la *conoscenza delle leggi naturali della popolazione*. In breve la Statistica, come scienza, sarebbe per tali scrittori quella che da altri con maggiore proprietà e chiarezza si chiama « Demografia » o « Demologia ». A questi altri noi facciamo l'adesione più completa e convinta.

Invero non ci par dubbia la convenienza di distinguere, fin nella nomenclatura, ciò che è sistema di cognizioni in un dato àmbito di fatti (*scienza*), da ciò che è strumento logico o regola di condotta per arrivare a tali cognizioni (*metodo*). Appunto perchè semplice norma di condotta del pensiero, uno stesso metodo può servire a diverse discipline; e una stessa disciplina, appunto perchè sistema di verità razionali, può giovarsi di parecchi metodi. Or come si giustificherebbe l'identità di nome, dove mancano l'identità di natura e l'identità di estensione? Il metodo statistico non è una specialità degli scienziati, che s'affaticano intorno alle leggi della popolazione; il fisiologo il quale osserva in che misura certi animali sono fecondi o in che proporzioni nascono maschi o femmine; il meteorologo che registra le temperature, le precipitazioni acquee, le variazioni diurne o annue della pressione atmosferica, ecc.; l'economista che mette a confronto

le variazioni dei prezzi e quelle della produzione dei metalli preziosi, l'alienista che numera i casi in cui una data forma di pazzia si rivela ereditaria, ecc., non fanno che applicare il metodo statistico a qualche parte delle loro discipline. Adunque il metodo in parola, nelle sue applicazioni, sconfina molto dai limiti di quella scienza analitica degli aggregati sociali, alla quale vorrebbe legare il proprio nome (1).

Importa pure far cessare la confusione o promiscuità della Statistica metodologica colle *Descrizioni statistiche di fenomeni sociali*, queste descrizioni rientrando nella competenza propria di speciali discipline, come la Demografia, l'Economia politica, la Scienza delle finanze, la Scienza dell'amministrazione, ecc. Il volerne fare un organismo dottrinale, un corpo solo sotto il nome di « Statistica descrittiva » è voler abbracciare troppe cose in una volta; nè si potrebbe dire d'aver fatto opera scientifica, quando in appendice alla Statistica metodologica, la quale ha un programma ben determinato, si fosse aggiunta una illustrazione di materiali statistici necessariamente superficiale, perchè la mole di essi, già enorme oggi, va ogni giorno aumentando.

---

(1) Questo scrivemmo nei nostri *Principii di Demografia* (Manuali Barbèra, Firenze 1900, Introduzione, pag. 10, 11) e non abbiamo avuto sinora motivo di cambiare opinione.

CAPO TERZO

Partizione della materia.

Sommario: § 1. Capitoli principali della Statistica, considerata come forma di osservazione. — § 2. Capitoli principali della Statistica come forma di induzione. — § 3. Motivi per assegnare in quest'opera un posto affatto secondario alla storia della Statistica.

§ 1. Una partizione razionale della materia presenta non poche difficoltà. Vi sono argomenti, di cui bisogna presupporre una certa conoscenza, che ogni lettore può attingere al suo fondo di coltura generale, finchè non venga il loro turno di trattazione sistematica; e ve ne sono che, per certi aspetti, converrebbe far precedere, per certi altri far seguire allo svolgimento di un dato ordine di questioni. Per esempio, la teoria degli errori e quella dei limiti dell'osservazione implicano l'induzione; tuttavia non è dubbia l'opportunità, almeno didattica, di collocarle in sede di Statistica osservatrice. È ragionevole anche fare la critica dei dati subito dopo la rilevazione e prima di affinare i materiali greggi coi processi delle medie, dei rapporti, delle perequazioni, ecc.; ma sotto un altro rispetto converrebbe invece far precedere questo lavoro di affinamento, in quanto agevola la critica e la scoperta degli errori. Ad ogni modo, senza scostarci troppo dalla traccia comune degli autori, tratteremo nel libro I della *Statistica come forma di osservazione* e nel II della *Statistica come forma d'induzione*. In appendice sarà dato un *Cenno storico della Statistica*.

I capitoli principali del libro I avranno per oggetto:

La rilevazione dei dati e la formazione delle tabelle primitive;

La critica e comparazione dei dati in questo stadio;

I procedimenti semplificativi e rappresentativi delle quantità rilevate, distinti in:

a) *aritmetici* (medie, perequazioni, rapporti);

b) *geometrici e grafici* (diagrammi, cartogrammi);

c) *algebrici* (interpolazioni, correlazioni).

Gli elementi del calcolo combinatorio e delle probabilità, nelle loro applicazioni ai fenomeni collettivi;

La formazione delle tabelle derivate (in particolare dei metodi di costruzione delle tavole di mortalità e di sopravvivenza);

La tecnica e logica dei numeri-indici, considerati come espressioni sintetiche di situazioni complesse;

La comparazione dei dati elaborati, con particolare riguardo alle statistiche internazionali.

Questa parte, a cui faremo seguire a titolo di pratica illustrazione un *Cenno sui censimenti della popolazione* (massime dell'ultimo nostro), esaurisce il còmpito della Statistica come forma di osservazione meto-

dica. Vi abbiamo fatto entrare di proposito, da un lato, i ***procedimenti semplificativi*** e ***rappresentativi*** delle quantità rilevate, dall'altro i procedimenti ***sintetici*** dei numeri-indici, perchè, prima ancora di affrontare il problema delle cause, l'ideale dell'osservatore dev'essere quello di ridurre un ***maximum*** di materiale greggio ad un ***minimum*** di elementi caratteristici, su cui fissare l'attenzione. Per tale rispetto la ripugnanza, che è in molti, dall'impiego delle matematiche, non ha ragione d'essere. Ad essi tuttavia si può fare questa concessione (consigliata da motivi didattici facilmente spiegabili, se si pensa che la Statistica è materia d'insegnamento della Facoltà giuridica), di non oltrepassare i limiti della coltura matematica impartita dai nostri licei. Anche entro questi limiti, il calcolo serve a liberare il campo di studio da molte erbe parassite o inutili, mettendo in evidenza delle regolarità, somiglianze, concomitanze alle volte meravigliose e indicando nel modo più suggestivo la via alle successive induzioni.

§ **2.** Il libro II tratterà appunto della Statistica come forma d'induzione. Tratterà cioè dei *metodi di induzione sperimentale*, nei limiti in cui essi si chiariscono applicabili ai fenomeni collettivi. Noi speriamo di poter colmare qui una lacuna dei logici, ai quali è sfuggito il carattere particolare dell'induzione statistica. Dei ***procedimenti deduttivi***, che bene spesso non sono che induzioni dissimulate e ad ogni modo non ci sembrano la caratteristica della nostra disciplina, si discorrerà pure in questa sede, il che ci darà motivo a passare in rassegna le ***ipotesi*** più frequentemente usate nell'analisi dei fatti collettivi e ricordarne l'applicazione nella ***rilevazione congetturale***.

3. Chiuderà l'opera un *Cenno storico della Statistica*, succinto quanto occorre per trovarci in accordo col nostro modo di vedere, che è questo: La storia di una scienza non è la scienza stessa; può essere un ordine di cognizioni estremamente interessante intorno allo svolgersi del pensiero umano in questa o quella direzione, ma non fa parte necessaria dei fatti, attorno ai quali il pensiero si è esercitato e seguita ad esercitarsi. Così non è indispensabile la storia delle matematiche alla dimostrazione dei teoremi di matematica; non quella della chimica per la dimostrazione dei fenomeni chimici, ecc. In un trattato, adunque, il quale abbia per oggetto una scienza determinata, un sistema di verità razionali, la storia di questa scienza o di questo sistema non può pretendere che ad un posto affatto secondario, ad un capitolo di introduzione o di chiusa; come in un trattato, il quale abbia per oggetto la storia di una scienza, le dimostrazioni proprie di questa non debbono figurare che incidentalmente, quando ciò sia voluto da particolari opportunità dell'esposizione. La divisione del lavoro e delle attribuzioni fa regola anche a questo riguardo.

PARTE SPECIALE

LIBRO I.

LA STATISTICA COME FORMA DI OSSERVAZIONE

CAPO PRIMO

La rilevazione dei dati e la formazione delle tabelle primitive.

TITOLO I.

LA RILEVAZIONE DEI DATI

Sommario: § 1. Piano della rilevazione e determinazione dell'oggetto. — § 2. Limiti di quantità dei casi da osservarsi, limiti di tempo, di luogo, di specializzazione, di precisione. — § 3. Se questi limiti formino sistema. —§ 4. Chi può fare la rilevazione. — § 5. Forme e modi della rilevazione. — § 6. Mezzi e strumenti. — § 7. Metodo e modelli di rilevazione di alcune statistiche italiane.

§ **1.** La *rilevazione dei dati* ossia « il complesso delle operazioni necessarie per venire a conoscenza dei casi di un fenomeno collettivo e per enumerarli nelle classi e graduazioni assegnate » suppone prestabilito un piano o programma di lavoro.

Il piano o programma deve comprendere la determinazione precisa dell'oggetto, quella dei limiti di tempo, di spazio, di specializzazione e di precisione dell'indagine, quella dei modi e mezzi di esecuzione e infine la previsione e prevenzione delle cause di errore.

Si capisce che i diversi punti del programma formano un sistema e non sono indipendenti l'uno dall'altro. Prestabilito un certo modo di esecuzione, non saranno più del tutto arbitrari i limiti di spazio e di tempo o di specializzazione che avrà l'indagine, e le cause temibili di errore non saranno le medesime che si avrebbero adottando altro metodo; del pari, prestabilite certe condizioni di tempo, di spazio, ecc.,

ne risulterà più o meno vincolata la scelta dei mezzi e modi. Se, ad es., decido di eseguire da solo, senza collaborazione altrui, una certa indagine, è ovvio che questa non andrà oltre il tempo utile della mia vita, nè al di là di un certo territorio; se adotto un questionario, al quale gli interrogati siano liberi di rispondere, o non rispondere, vuol dire ch'io mi contento anticipatamente di quel qualunque numero di casi che risulterà dalle risposte volontarie. Ora certamente esiste per ogni ricerca una combinazione di elementi, modi e mezzi, tale da realizzare il principio del massimo risultato utile col minimo di costo; ma il trovarla è cosa la quale dipende più dall'intuito dell'osservatore che dalla disciplina di regole metodologiche.

Prima cura intanto dev'essere di ben definire l'oggetto dell'investigazione, che è quanto dire fissare quei caratteri che, mentre sono comuni a tutti i casi della classe considerata, li differenziano tutti e ciascuno dai casi di classi anche affini. Sia, per es., da accertare in un paese la frequenza della *nati-mortalità*. Bisognerà por mente a che non si denuncino come nati-morti dei semplici aborti e nemmeno dei nati vivi e non vitali e molto meno ancora dei nati vivi e vitali, ma decessi per accidentalità qualche minuto o qualche ora dopo la nascita. I caratteri comuni a tutti i casi della classe « nati-morti » e che ad un tempo li differenziano dalle classi affini nominate, saranno: 1° lo sviluppo del feto, giunto, come si dice, a termine; 2° il mancato atto della respirazione al momento della nascita. Il primo di detti caratteri varrà a distinguere i nati-morti dagli aborti; il secondo li distinguerà dai nati-vivi, vitali o non vitali che siano. Certo, trattandosi di fenomeni che differiscono per gradi, possono sussistere ancora cause di errori o motivi di diverso apprezzamento: così, è difficile indicare se il feto è giunto o non è giunto a termine; nè la difficoltà si elimina del tutto fissando un dato numero di mesi (per esempio, *sei*) dal concepimento.

Alcuni osservatorii meteorologici considerano come giorni piovosi anche quelli in cui non vi fu vera pioggia, ma solo condensazione di umidità nel pluviometro; altri invece, più correttamente, tengono conto dei soli giorni in cui si ebbe *precipitazione*. Caratteristica differenziale della pioggia è appunto la precipitazione per forza di gravità del vapor acqueo condensato in forma di goccie d'acqua, inette a galleggiare nell'atmosfera.

La domanda della religione nel nostro ultimo censimento era formulata in modo da lasciare il dubbio in molti, se fosse diretta ad accertare uno stato di fede, di coscienza, indipendentemente dal fatto della nascita o l'appartenenza per nascita ad un dato culto, indipendentemente dalle intime convinzioni. La domanda era: « *Religione. Chi appartiene ad un culto, dica qual è* ». Secondo la prima interpretazione, un cattolico o un israelita per nascita potè rispondere ben anche: *ateo*, *razionalista*, *libero pensatore*, ecc.; secondo l'altra e più corretta, gli stessi atei, razionalisti, ecc., poterono denunziarsi catto-

lici, israeliti od altro. E deve dirsi corretta la seconda interpretazione, non essendo presumibile, in un regime di riconosciuta libertà di coscienza, che lo Stato volesse permettersi inchieste, con carattere imperativo, sulla fede religiosa o politica; lo Stato non poteva pretendere se non le notizie di fatti esteriori e già di dominio pubblico od altrimenti risultanti da documenti, come gli atti di battesimo e simili.

Quindi l'oggetto della rilevazione vuol essere certo, concreto e ben determinato per tutti quei caratteri che decidono dell'omogeneità della massa e della sua netta separazione da masse anche affini.

Già sappiamo che fra un caso e l'altro della stessa specie esistono differenze di grado o anche vere discontinuità. Havvi discontinuità tra un sesso e l'altro, tra la qualità di cittadino e quella di straniero, tra il mezzo « arma da fuoco » e il mezzo « veleno » o « annegamento » prescelto da un suicida, e così via; cioè non si passa per gradi insensibili da una forma all'altra di manifestazione del fenomeno. Havvi invece possibilità di gradazioni quante si vogliono fra un'età e le altre età, fra una data statura e le altre stature, ecc. Orbene nella determinazione dell'oggetto, che si intende di rilevare, regola generale è questa: le discontinuità e le differenze di grado, che passano fra un caso e l'altro della stessa categoria, debbono essere o reputarsi, agli scopi dell'indagine, meno importanti di quelle che separano una categoria da altra anche affine. La quale regola ammette pure questa forma di enunciato: *Tra uno dei casi estremi di una classe e uno degli estremi di classi confinanti ci dev' essere almeno una caratteristica differenziale in più, che fra i due estremi della stessa classe* (1).

L'importanza di questa norma non ha bisogno di essere dimostrata. Gli errori qualitativi, ossia gli errori che derivano da una imperfetta determinazione dell'oggetto dell'indagine, si traducono in errori di quantità, vuoi per l'inclusione di casi estranei alla classe considerata, vuoi per l'esclusione di casi, che invece ne dovrebbero far parte.

Definito l'oggetto della rilevazione, ogni singolo caso direttamente osservato, o comunque venuto a conoscenza dell'organo raccoglitore, costituisce una *unità statistica*.

§ **2.** L'enumerazione, la classificazione e graduazione, la disposizione in serie di tempi, assai facili quando i casi sono pochi e con caratteri non dubbi, richiedono operazioni vaste, complesse e delicate, così nel predisporre i mezzi e i modi per la raccolta delle notizie,

---

(1) Supponiamo che i casi estremi della specie « *suicidii compiuti* » si trovino separati da una differenza di sesso, da una di età, da una di stato civile, da una differenza di mezzo, ecc. Se ora si prende uno degli estremi di questa specie e lo si confronta coll'estremo opposto della specie confinante « *suicidii tentati* » è chiaro che avremo, oltre le differenze sopradette, anche quella dell'*esito dell'atto*, che nell'un caso è *compiuto* e nell'altro *interrotto*.

come nell'eseguire gli spogli e gli aggruppamenti più razionali, allorchè i casi si presentano a migliaia e a milioni e si vogliono classificare per elementi combinati anzichè isolati. I recenti censimenti della popolazione in Russia e nell'India inglese, l'inchiesta antropometrica americana al tempo della guerra di secessione, e quella italiana sugli arruolati delle leve 1879-1883, i censimenti professionali in Germania, forniscono esempi di siffatte colossali operazioni statistiche. Non senza qualche fondamento il Bowley definiva testè la statistica *The science of counting* (1), beninteso se si prescinde dal maggior còmpito che essa ha come forma d'induzione. In ragione appunto delle difficoltà che s'incontrano, è illusorio fidare in una esattezza àssoluta; bisogna starsene paghi di una buona approssimazione. Ciò d'altronde avviene anche nelle scienze fisiche; si può spingere nel calcolo dei coefficienti di dilatazione lineare dei solidi l'approssimazione fino al centomillesimo di millimetro; ma ci si può contentare di conoscere la distanza della luna o del sole dalla terra coll'approssimazione di parecchie decine o parecchie migliaia di leghe, rispettivamente. Così non si accetterà come rigorosa la cifra di 60 metri al minuto secondo per la velocità di propagazione della corrente nervosa nelle fibre motrici, nè quella di 130 metri nei nervi sensitivi dell'uomo. Nei fenomeni collettivi sociali si va dalle quasi esatte registrazioni delle nascite, delle morti, dei matrimoni, a quelle assai imperfette delle produzioni agrarie, dei redditi, ecc., in cui l'errore temibile può eccedere la misura del 10 o del 20 per cento. Della qual cosa bisogna tener conto (come vedremo parlando della tecnica e logica dei numeri-indici) allorchè si tratti di confrontare e comporre in una le variazioni di fenomeni diversi riflettenti i diversi aspetti di una situazione complessa di cose: le singole serie debbono in tal caso entrare in calcolo, per ragione di omogeneità, con un peso o coefficiente d'importanza in ragione inversa del loro grado di certezza statistica.

I fenomeni collettivi si svolgono intorno a noi in modo incessante, ma a volerli seguire tutti e sempre, nelle loro variazioni infinitesimali e nelle loro complicatissime combinazioni, occorrerebbe un lavoro infinito. Per renderlo finito e possibile, si suggerisce di limitare o i casi d'osservazione o il tempo o lo spazio o la specializzazione delle classi o la precisione. Questi diversi limiti, anche combinati tra loro, hanno conseguenze, che presto vedremo; essi ci sono in ultima analisi imposti da considerazioni riguardanti il costo e l'utilità della ricerca, la probabilità di risultati più o meno attendibili, i diritti, gli interessi, i pregiudizi e le resistenze degli interrogati, l'urgenza medesima che possiamo avere di conoscere i risultati, massime quando l'indagine è intrapresa per fini amministrativi o politici.

---

(1) Arthur L. Bowley, *Elements of Statistics*, pag. 3, 2ª edizione, London, P. S. King, 1902.

Una teoria completa dei limiti della rilevazione è ancora oggi un *desideratum* della metodologia statistica; un primo passo per arrivarci facciamolo raccogliendo sparsi frammenti.

a) *Limitazione dei casi di osservazione.* — Invece di rilevare tutti i casi possibili, possiamo prenderne un certo numero *a sorte*, ovvero sceglierne meditatamente alcuni *quasi tipici*. Se la limitazione si opera mediante scelta di casi quasi-tipici (l'assolutamente tipico non è oggetto di osservazione statistica o per masse), i criteri per la scelta migliore debbono domandarsi a tutte quelle cognizioni, che già possediamo intorno al fatto, le quali tendono a rappresentarcelo come il caso medio o di maggior frequenza. Così, volendo determinare il bilancio domestico, entrate e spese, di un certo numero di famiglie operaie, sceglieremo di preferenza le famiglie che si possono considerare ad un tempo tipiche, nel significato discreto della parola, per numero di componenti (per es. 5 o 6, ma non 1 o 2 soltanto e nemmeno 10 o 12), per entità di salario (nè troppo basso, nè troppo alto e soprattutto non integrato da considerevoli redditi di capitale), per l'attività attuale d'impiego, almeno del capo-famiglia (esclusi dunque i casi estremi di disoccupazione prolungata o i periodi di straordinario lavoro), ecc. Il che non toglie che si possa fare una punta anche nel campo delle eccezioni od anomalie.

Se la limitazione si opera a sorte, è necessario che i rilievi siano abbastanza numerosi per escludere l'influenza di cause accidentali e conferire attendibilità alle nostre generalizzazioni. Esempio ben noto è quello della verifica del peso e del titolo delle monete, che la Commissione di controllo della circolazione in Francia va compiendo da alcuni anni. Nel 1901 essa esaminò 41.828 pezzi di monete di un millesimo qualunque, presi dalla circolazione, classificandoli secondo che eccedevano per peso (monete forti) o stavano nei limiti della tolleranza di fabbricazione, oppure stavano nei limiti della tolleranza di logoro o erano inferiori a detti limiti (monete leggiere). Allorchè un rilievo si fa diramando questionari, cui gli interrogati siano liberi di rispondere o non rispondere, si ha limitazione a sorte, eventualmente combinata con altre limitazioni.

b) *Limiti di tempo.* — Bisogna distinguere tra fenomeni di *stato*, cioè complessi di casi che coesistono in un medesimo punto del tempo, e enomeni di *movimento*, cioè complessi di casi che si succedono durante un intervallo, modificando la situazione iniziale. La popolazione, per esempio, a un istante dato, è una coesistenza di individui o fenomeno di stato; le nascite, le morti, le migrazioni, che di continuo la fanno variare, sono fenomeni di movimento. Uno stato o una situazione esprimono le cambiali in portafoglio a fine mese presso una Banca; ma i nuovi effetti scontati che entreranno in portafoglio giorno per giorno e quelli che ne usciranno perchè pagati alla scadenza, sono le varianti della situazione, i fenomeni di movimento.

Ora i casi di una coesistenza possono avere una vita più o meno lunga e un movimento di rinnovazione e di estinzione più o meno regolare. La vita media di una generazione si calcola di 35 o 36 anni; la scadenza media degli effetti scontati da un Istituto di credito sarà, poniamo, di 3 o 4 mesi. Un debito ipotecario si contrae in media per la durata di 10 o 12 anni, ecc. Gli è evidente che quanto più duraturi sono i casi, col ripetere a brevi intervalli i rilievi della loro situazione noi non facciamo che riosservare e ricontare molti degli stessi casi già osservati e contati altre volte. Si affaccia quindi subito il partito di economizzare sul tempo, rinnovando l'osservazione solo a periodi, in capo ai quali un mutamento apprezzabile si possa credere avvenuto. L'intervallo sarà breve per i fenomeni di breve durata e di movimento irregolare; più o meno lungo per gli altri. Del portafoglio di una Banca converrà, poniamo, accertare le situazioni decadarie o settimanali; ma per una popolazione il censimento può ripetersi senza inconvenienti solo ogni decennio.

Se di un fenomeno fosse conosciuto con esattezza matematica tutto il movimento, la situazione, una volta accertata, si potrebbe dire accertata per sempre. Il calcolo dispenserebbe da ulteriori rilievi generali, bastando aggiungere volta per volta alla situazione iniziale i nuovi casi, che costituiscono in certo modo l' « entrata » e togliere quelli che costituiscono l' «uscita». Spesso però, anzi sempre, il movimento è conosciuto solo in parte o in maniera imperfetta e approssimativa. Siamo, ad esempio, bene informati intorno alle nascite e alle morti, male intorno alle immigrazioni, ai rimpatrii e alle emigrazioni; sicchè la revisione periodica del numero degli abitanti si rende necessaria. In Italia la situazione del debito ipotecario non è stata più accertata dopo il 1871. Ora il movimento del debito ipotecario è così imperfettamente determinabile, per cause delle quali avremo occasione di parlare, che dieci o dodici anni dopo quella data non v'era certo più corrispondenza fra la situazione calcolata e la reale. Converrebbe appunto da noi stabilire un periodo di revisione non maggiore di dodici o quindici anni, tenuto conto che questa è a un bel circa la scadenza media di tali debiti. Generalizzando dunque diremo che il grado di limitazione delle nostre conoscenze relative al movimento di un fenomeno, decide della lunghezza d'intervallo che convien dare a due successive verifiche dello stato del fenomeno stesso.

La rilevazione di un fenomeno di movimento continuo è generalmente continua; ma allorquando il tempo costituisce un fattore di variazioni assai meno importante di altri fattori, ragioni di economia possono consigliare qui pure una limitazione. I casi in cui una statistica continuata per uno o più anni viene interrotta sono troppo comuni perchè si abbia bisogno di citarli.

Qualche volta l'osservazione si restringe a tempi, per dir così, tipici, che presentano gli estremi del fenomeno, mentre nell'intervallo di

tempo che separa un estremo dall'altro il fenomeno non fa che passare per gradazioni continue dall'una all'altra forma tipica. Ell'è, sotto apparenza un poco mutata, ancora una limitazione per scelta di casi quasi-tipici. Così per uno studio sulle variazioni diurne della temperatura o della pressione barometrica potremo contentarci delle registrazioni limitate ai 20 o 25 giorni circostanti al solstizio d'estate e a quello d'inverno, le due epoche più caratteristiche dell'anno.

c) Combinazione dei limiti di quantità e di tempo. — Questa combinazione presenta un particolare interesse in molte ricerche. Per conoscere, ad es., la legge di crescimento dell'uomo in statura o in peso, non è necessario registrare giorno per giorno la statura o il peso di migliaia di persone, seguendole nominativamente nel corso della loro vita; basterà tener nota di un piccolo numero di soggetti. Se si desidera calcolare su masse considerevoli, bisognerà limitare la frequenza delle registrazioni — una volta per stagione o per anno. Ma anche questo procedere ci obbligherebbe (se si seguono nominativamente i soggetti d'osservazione nel corso della loro vita) ad aspettare un secolo per venire a capo dell'intera serie. Allora si gira la difficoltà, combinando i limiti di quantità e di tempo nel seguente modo: si misura la statura (o il peso od altro carattere) di 1000 bambini alla nascita, di 1000 *altri* bambini di 1 anno coesistenti coi primi, di 1000 *altri* bambini di 2 anni d'età, pure coesistenti coi primi e coi secondi, e così via; presumendosi che gli attuali neonati avranno fra dodici mesi la stessa statura media degli attuali bambini di 1 anno, e fra due, tre, dieci anni, la stessa statura media degli attuali fanciulli che hanno compiuto il secondo, il terzo, il decimo anno d'età. La qual presunzione regge (per tacere di altre riserve), finchè non sia provato che alla mortalità infantile contribuiscono relativamente di più i nati di bassa statura, come gli atrofici, assai frequenti tra i gemelli, gli illegittimi e i primogeniti di madri precoci; poichè in tal caso i sopravviventi alle varie età deriverebbero da una categoria scelta di nati, in cui preponderano i meglio conformati aventi una propria media superiore a quella ottenuta su 1000 neonati indistintamente presi.

d) Limiti di spazio. — La limitazione di spazio si opera generalmente mediante scelta di località riconosciute tipiche per la produzione di un dato fenomeno, trascurando le altre o non specializzando in queste le indagini allo stesso segno che in quelle. Così possiamo studiare statisticamente una zona specifica di malaria, un gruppo etnico nei distretti dove ha una decisa prevalenza, i prezzi di una merce nelle località di maggior produzione o di maggior consumo, ecc.

e) Combinazione dei limiti di quantità e di spazio. — È bene tradotta in pratica dalle così dette « investigazioni rappresentative » (1),

---

(1) KIAER, *Observations et expériences concernant des dénombrements représentatifs* (*Bull. de l'Institut internat. de Statistique*, vol. IX, fasc. 2º. Roma).

raccomandate dal Kiaer. Si tratta, dice il Bosco, di esplorazioni parziali limitate ad un certo numero di località sparse per il paese, distribuite e scelte in guisa da riprodurre la fisionomia di questo. Così, per rilevare in Norvegia il reddito e la ricchezza delle varie classi sociali « furono trascelti per tutto il territorio un certo numero di Comuni e di città, in modo da corrispondere alla proporzione, secondo la quale la popolazione rurale e la urbana si ripartiscono nello Stato; e fu del pari fissato per ciascun Comune o città un certo numero di abitanti; rispetto ai quali soltanto dovevano essere raccolte le notizie. Che i criteri seguiti fossero giusti, fu dimostrato dal risultato finale: fra i cittadini, rispetto ai quali si esercitò l'osservazione statistica, le professioni si trovarono distribuite in modo analogo a quello indicato dal censimento per tutta la popolazione. La piccola parte di abitanti, a cui fu limitata l'indagine, rendendo così l'immagine della intiera popolazione, si potè dare un valore generale ai fatti rilevati intorno all'ammontare e alla ripartizione del reddito ». Il Bosco nota giustamente che questo metodo, il quale si risolve in una limitazione combinata di casi e di luoghi e si chiarisce applicabile in certe indagini demografiche ed economiche, non andrebbe accolto, per esempio, nella statistica penale. « Non si può dalle condizioni criminali di pochi luoghi, anche opportunamente scelti, dedurre quelle di un intero paese, giacchè i delitti variano troppo di numero e di specie nelle varie regioni, per effetto delle cause locali, che si sovrappongono e modificano l'azione delle cause generali » (1). Qui si ha piuttosto bisogno di raccogliere la maggiore quantità possibile di casi e per ogni punto del territorio, che non di limitarne il numero.

f) Limiti di specializzazione. — Come già abbiamo detto, i casi di un fenomeno collettivo sono suscettivi, grazie ai loro molteplici caratteri, di svariatissimi aggruppamenti, perchè non esiste caso perfettamente identico ad un altro. Ma non potendosi andare all'infinito, converrà arrestare la specializzazione agli elementi più importanti. L'importanza è, a dir vero, un concetto soggettivo, cui deve corrispondere in ultima analisi alcunchè di oggettivo, fuori di noi, come proprietà delle cose osservate o come loro relazioni con altre cose. Noi conosciamo le cose per le loro somiglianze e differenze; la relatività delle nostre cognizioni è un dato fondamentale della Logica. In Statistica andiamo appunto in cerca di contrasti. Le stesse somiglianze non sono che contrasti di caratteri particolari, associati a identità generali. Una classificazione, che non mettesse in evidenza un contrasto, si potrebbe lasciare in disparte: ma a tutto rigore questo caso non si verifica mai. Certo però è che, a parità d'altre circostanze, fra due modi di classificazione da cui si aspettano due contrasti di situazioni inegualmente accen-

(1) A. Bosco, *La Statistica civile e penale e la riunione dell'Istituto internazionale di Statistica a Pietroburgo*, pag. 61 e 62, Roma, Bertero, 1898.

tuati, richiama primo la nostra attenzione quello del contrasto più accentuato.

Supponiamo di dover fare la statistica delle « separazioni coniugali ». Distingueremo le istanze di separazione secondo la *provenienza* (dal solo marito, dalla sola moglie o da entrambi i coniugi, sia per mutuo consenso, sia per atto separato, anche in via riconvenzionale); secondo l'*esito* (che può essere di desistenza o riconciliazione, di rigetto o di accoglimento); secondo i *motivi* pei quali venne accordata la separazione (condanna penale, adulterio, eccessi, sevizie e minaccie gravi, ecc.); secondo una moltitudine di *circostanze personali* dei coniugi (età assoluta o relativa, durata della convivenza trascorsa, esistenza o meno di figli, condizione economica, ecc.). E non è tutto. Potremmo classificare ancora, combinando questi ed altri elementi due a due, tre a tre, ecc. Conseguenza di ciò è che si discoprono sempre nuovi contrasti di situazioni e si va incontro ai piccoli numeri.

Ora, se c'è ragion di credere, ad esempio, che la mancanza di figli implichi una frequenza relativa di separazioni coniugali assai diversa da quella che si ha nel caso d'esistenza di figli, laddove l'essere o non essere il marito più vecchio della moglie non paia spiegare un'influenza altrettanto decisiva, è ovvio che noi sacrificheremo al bisogno la seconda distinzione alla prima, e non viceversa. In altri termini, prevarrà il criterio del contrasto più accentuato. Se si affaccia la duplice indagine delle separazioni secondo lo stato economico, ricco, agiato, medio e povero, dei coniugi, e secondo le professioni, e d'altronde l'economia del programma imponga di optare per l'una o per l'altra, sceglieremo la prima, come quella che ci presenta il caso di una correlazione più immediata. Infatti è a ritenersi che la proclività alle separazioni sia funzione piuttosto del tenor di vita concomitante ai diversi gradi di agiatezza, che non del genere di professione esercitata; la professione c'entra solo indirettamente, cioè in quanto si accompagna ad un reddito più o meno elevato. Naturalmente, quel che abbiam detto vale quando si tratti di rilevazione *ex novo*, senz'altro precedente che l'osservazione comune; perocchè, una volta assicurata con rilievi regolari la conoscenza delle cause, che dànno luogo ai maggiori contrasti di situazioni, si può abbandonare per qualche tempo l'indagine in questa direzione, per rivolgerla ad altri elementi, meno caratteristici forse, ma anche meno esplorati.

Una conseguenza inevitabile dello specializzare è poi l'esiguità numerica dei gruppi e ben anche il loro isolamento da altri gruppi per mezzo di classi vuote, di lacune. Criteri assoluti per dire dove finiscono i grandi numeri e incominciano i piccoli, non si conoscono; bisogna rimettersene al « senso statistico » di chi fa la rilevazione. Ci sono dei numeri di poche unità e tuttavia molto grandi in confronto di quelli che si avrebbero in loro luogo e vece, se le combinazioni dei fattori, da cui derivano, avvenissero come in un giuoco di sorte, giusta i principii

delle probabilità. Esempio: a Milano si conta ogni anno una ventina di matrimoni tra lavandai e lavandaie; a calcolo di probabilità non se ne dovrebbe avere che *uno ogni tre anni*. La cifra osservata, benchè piccola per se stessa, è 60 volte giù grande della calcolata e questo contrasto interessa lo statistico come dimostrazione dell' influenza che la somiglianza di mestiere esercita sulla scelta matrimoniale. Ci sono dei numeri piccoli che attestano la grande efficacia di una causa ostacolante; esempio: i rari matrimoni fra ebrei e non ebrei nei paesi ov'è diffuso l'antisemitismo. Ci sono infine dei numeri piccoli, che si riferiscono a gruppi sceltissimi, ottenuti per combinazione di parecchi caratteri ad un tempo. In quest'ultimo caso, specialmente, avviene che al di là di un certo limite l'esiguità dei gruppi non consente più la determinazione sicura dell'elemento tipico e lascia sussistere il dubbio che le cause accidentali abbiano avuto buon giuoco sulle costanti. A questo punto, che l'occhio statistico bene esercitato sa discernere, si deve arrestare la specializzazione. Lo stesso si dica quando, essendoci motivo di credere che i casi del fenomeno in esame formino un sistema continuo, la specializzazione operata in un campo d'osservazione ristretto dia origine a lacune, a classi vuote, non solo ai confini, ma nel mezzo stesso del quadro di classificazione, tra un gruppo e l'altro; infatti noi siamo allora obbligati, almeno per comodità di studio, a restituire al sistema la sua presunta continuità, facendo un passo addietro nella specializzazione, cioè fondendo insieme classi contigue e adottando più larghe graduatorie.

Subordinatamente dunque ai limiti generali imposti dal costo e dall'utilità della ricerca, dalla complicatezza dei questionari, ecc., si conclude:

1° la specializzazione non deve andare oltre il punto al quale la maggioranza dei gruppi consterebbe di piccoli numeri inetti alla determinazione di valori tipici sicuri e presenterebbe molti vuoti o soluzioni di continuità;

2° conviene specializzare, ma solo in quelle direzioni in cui, per assaggi o per osservazione comune, si prevede maggiore il contrasto dei particolari;

3° gli elementi di correlazione più diretta debbonsi preferire a quelli di correlazione indiretta.

Ma, ripetiamo, in tutto ciò vi ha poco di assoluto; l' intuito dell'osservatore e la sua attitudine a valutare le circostanze del caso valgono spesso meglio di ogni regola formale.

g) Limiti di precisione. — Se la conoscenza esatta di un fenomeno collettivo assai esteso implica difficoltà e spese sproporzionate all'utilità della ricerca, possiamo surrogarla con una conoscenza approssimata. Questa si fonda alle volte su *stime di persone pratiche* (come nelle statistiche agrarie), su *denuncie di interessati* non rigorosamente controllate (statistiche dei redditi di ricchezza mobile, delle successioni

ereditarie e simili), su *misure eseguite con strumenti imperfetti*, ma di uso rapido (come i rilievi sull'indice cefalico fatti mediante il « quadro a massima » anzichè mediante il « compasso »), su *indicazioni generiche* invece che specifiche, in termini *qualitativi* invece che quantitativi (come la classificazione di tessuti finissimi, fini, medi, ordinari, in luogo di tessuti di un dato peso, per 100 metri quadrati contenenti tanti o tanti altri fili fra catena e trama per centimetro quadrato, in una statistica doganale), su *enumerazioni*, dalle quali non si possano o per speciali difficoltà non si vogliano eliminare i *casi duplicati* (come nelle statistiche dei proprietari di immobili, desunte dai ruoli delle singole Agenzie d'imposte dirette, senza riscontro nominativo delle iscrizioni ripetute allo stesso nome in Agenzie diverse), e così via discorrendo.

§ **3.** I diversi limiti, di cui abbiamo trattato, formano certamente sistema. Si può sacrificare la precisione alla continuità, la continuità all'estensione nel territorio, ecc., ovvero si può limitare da più parti ad un tempo la rilevazione. Havvi regola generale assegnabile per ottenere un *maximum* di effetto utile o un *minimum* di perdita? Come in un sistema di soddisfazioni economiche si dimostra che un massimo edonistico viene assicurato quando i diversi piaceri sono portati ad un egual grado finale d'intensità o sazietà, così nell'ordine delle soddisfazioni intellettuali, che derivano dalle conoscenze di fenomeni collettivi, un massimo di vantaggio relativo si avrebbe spingendo le indagini, nelle diverse direzioni che esse posson prendere, sino ad egual grado finale di certezza o completezza. Perchè infatti conseguenza di ogni limitazione è una diminuzione di certezza. Questo s'intende immediatamente quando si tratti di sostituire una rilevazione approssimata ad una rigorosa; ma s'intende pure quando restringiamo a cento o a mille casi una rilevazione che potrebbe comprenderne centomila o un milione, e quando separiamo in poche classi una massa di elementi che sarebbe suscettiva di una conoscenza assai più specializzata. Ma il teorico non può andar molto in là della regola sopra indicata, la quale ha d'altronde un valore di mera analogia; tocca all'uomo pratico trovare, anche per tentativi, il sistema di limiti più conveniente, tenendo conto degli interessi scientifici o amministrativi impegnati nella ricerca, della spesa, degli ostacoli probabili e d'altre circostanze ancora. Gli è perciò che assume una particolare importanza la scelta delle persone che s'incaricano di dirigere operazioni statistiche.

§ **4.** *Chi* può fare una rilevazione? Chiunque, un privato, un'associazione privata, un ente pubblico. L'organo raccoglitore, quand'è collettivo, ha il vantaggio di una maggiore autorità, continuità, simultaneità d'azione su molti punti del territorio; dispone solitamente di grandi mezzi; può colla giustificazione di un pubblico interesse richie-

dere in forma obbligatoria le necessarie notizie a coloro cui si rivolge; si presuppone più imparziale nella condotta dell'indagine e nella pubblicazione dei risultati, la parzialità e i secondi fini non potendo a lungo rimanere protetti dal segreto per parte di tutte le persone che compongono l'organo collettivo. D'altro canto il privato, che abbia qualità eminenti di osservatore, spiega una maggiore oculatezza e diligenza in ragione dell'interesse scientifico, che per lui può avere una data conoscenza, adatta i mezzi alle speciali difficoltà del caso, e se l'osservazione statistica s'intreccia colla sperimentale, come talvolta avviene (per esempio, nelle indagini concernenti la selezione artificiale e gli incroci), la sua competenza non può essere sostituita con vantaggio da quella di un organo pubblico. Il difficile per il privato è di rimuovere il dubbio, ch'egli abbia tenuto nota solo delle osservazioni che venivano in appoggio ad una sua tesi preconcetta. Tutto sommato, si capisce che vi sono iniziative e sfere di competenza più adatte all'inchiesta pubblica che alla privata, e viceversa; ma una linea di divisione netta non si può tracciare. Modello d'indagini statistiche d'iniziativa privata sono quelle dell'inglese Galton sui caratteri ereditari delle famiglie (*Family records*).

Qualunque sia la persona che si assume una rilevazione statistica, il programma preparato può richiedere un vario numero, talora grandissimo, di esecutori, fra i quali vien diviso il lavoro di raccolta, di spoglio e aggruppamento e d'intavolazione dei dati; e se le notizie sono provocate per mezzo di questionari diramati al pubblico, in forma libera od obbligatoria, anche il genere di cooperazione, che può aspettarsi dal pubblico, costituisce materia degna di molta attenzione. Vogliam dire che il carattere, la coltura, l'intelligenza, lo spirito di disciplina, ecc., di un popolo sono pure elementi i quali decidono dei limiti da assegnare ai rilievi statistici. Ad un popolo, che è o si crede perseguitato dal fisco, è pressochè vano rivolgere domande allusive alla condizione economica delle famiglie; i più troveranno modo di non rispondere o di rispondere alterando il vero. Ad un popolo in gran parte analfabeta, non si debbono proporre se non questioni di una chiarezza e semplicità estrema, evitando soprattutto quelle che implicano apprezzamenti soggettivi, variabili da individuo ad individuo. Nell'ultimo nostro censimento erano di tal natura i quesiti della presenza occasionale e della assenza temporanea; e l'averli poi abbinati sulla stessa scheda con riserva che all'uno rispondesse l'assente, nel luogo di dimora occasionale, all'altro il capo-famiglia dell'assente, nel luogo di dimora abituale, aggravò l'inconveniente delle interpretazioni contraddittorie.

Nei rilievi estesi a tutto il territorio dello Stato e che per la loro stessa vastità non possono ordinariamente attuarsi se non a forma di pubblica funzione, hanno una speciale importanza la *scelta del personale dirigente ed esecutivo* e la determinazione delle *sfere di*

competenza dei diversi organi centrali e locali. Sono però questioni di cui l'esperienza addita caso per caso la miglior soluzione o che, in mancanza di pratica già formata, si risolvono per tentativi; *a priori* non si possono enunciare se non regole assai vaghe. Nessun dubbio che occorra un personale intelligente, zelante e disinteressato quanto ai risultati dell'operazione statistica, il quale proceda con uniformità di criteri, secondo la traccia indicata dal programma; la mala fede di quei pochi che, per risparmiar lavoro, inventano le cifre o le gonfiano per vantare benemerenze, dovrebbe trovare in un *sistema di improvvise verifiche* il suo correttivo. Ciò è necessario soprattutto quando la scelta del personale è in qualche modo già vincolata. È difficile fare a meno dell'opera dei cancellieri nelle statistiche giudiziarie, dei maestri nelle statistiche scolastiche, degli impiegati comunali nelle statistiche del movimento dello stato civile, ecc. Quanto alle attribuzioni degli organi centrali e locali, un punto non controverso è questo: che il lavoro di coordinamento e di elaborazione dei risultati parziali (oltre quello, ben inteso, di preparazione del programma e delle istruzioni per l'intelligenza dei questionari) dev'essere riservato all'organo centrale; ma, caso per caso, si potrà riconoscere opportuno l'affidare agli organi locali qualche cosa più del semplice spoglio delle notizie contenute nei registri o della semplice distribuzione dei questionari e ritiro delle risposte; si potrà affidare, ad esempio, il lavoro di prima revisione delle risposte, il quale implica alle volte l'interpretazione di termini dialettali, che il personale del luogo è più atto a comprendere, o implica la conoscenza delle condizioni proprie degli interrogati. Del resto, molto dipende, oltre che dalla specie di organi locali, dalle esigenze particolari dell'inchiesta. Ai singoli osservatorii meteorologici, alle singole segreterie universitarie, alle succursali di una grande Banca rimetteremo senz'altro lo stesso lavoro di elaborazione, restringendo il còmpito dell'organo centrale al coordinamento dei risultati parziali; ma se si trattasse di fare il rilievo dei proprietari in base ai ruoli delle Agenzie delle imposte dirette e si volessero eliminare le intestazioni ripetute allo stesso nome, l'accentramento dello spoglio diverrebbe una necessità; quanto alle singole Agenzie, basterebbe che s'incaricassero di accertare ed eliminare le iscrizioni duplicate nei limiti del proprio distretto.

Una certa libertà d'azione si desidera per gli organi locali, come quelli che nello svolgere il programma assegnato possono, senza alterarne le linee essenziali, spingere l'inchiesta a particolarità interessanti dell'ambiente locale. In questa maniera, accanto alle statistiche governative, si sviluppano le statistiche dei Municipi, delle Camere di commercio e simili, che, per la ricchezza di particolari e la varietà dei fenomeni illustrati, costituiscono fonti di prim'ordine per gli studi amministrativi, demografici ed economici.

§ 5. Le forme e i modi della rilevazione si possono classificare così:

a) ***Rilevazione preliminare o d'assaggio* e *rilevazione definitiva.*** — Sempre, quando si tratti di rilievi *ex novo*, senz'altro precedente che l'osservazione comune, e spesso, quando si tratti di operazioni assai vaste e complesse, convien procedere ad una rilevazione preliminare o d'assaggio, sia allo scopo di avere una norma per la distribuzione delle schede, dei questionari, ecc. in numero sufficiente ed all'indirizzo esatto degli interrogati, sia allo scopo di fissare con più sicuro criterio i limiti dell'indagine, il preventivo della spesa, l'opportunità di certe classificazioni. Così una statistica industriale si preparerà procurando a mezzo dei Sindaci e delle Camere di commercio un elenco completo delle ditte, distinte secondo la sede e la specie d'industria esercitata; un censimento della popolazione, richiedendo dai Comuni il numero probabile delle famiglie e degli individui secondo le risultanze dei registri d'anagrafe e disponendo perchè i commessi facciano una visita preventiva nelle case a scopo d'informazioni sul numero delle persone che compongono le famiglie o le convivenze. Anche dopo raccolte le schede, se vi ha motivo di tentare una classificazione o combinazione di elementi, che non faceva parte del programma primitivo, si può fare un assaggio su un numero limitato di schede, prese a sorte, dopo di che si procede alla rilevazione definitiva.

b) ***Rilevazione automatica* e *rilevazione riflessa***, secondo che i casi sono portati a conoscenza degli uffici raccoglitori dagli stessi interessati od obbligati per disposizione di legge, ovvero sono richiesti volta per volta d'iniziativa dell'amministrazione pubblica o d'iniziativa di un privato individuo o di una privata associazione.

La registrazione del movimento dello stato civile (nascite, morti, matrimoni, legittimazioni) è automatica; l'inchiesta in corso sulle condizioni della classe operaia a Milano è riflessa. I censimenti sono rilevazioni riflesse.

c) ***Rilevazione continua*, *periodica* ed *occasionale***, secondo che registra i casi man mano che si producono e si succedono nel tempo, o invece li sorprende, per dir così, a momenti determinati e separati da intervalli regolari o prefissi, o infine li accerta solo quando circostanze speciali chiariscono l'opportunità dell'osservazione metodica. Del primo genere sarebbero le statistiche, già ricordate, del movimento dello stato civile, le statistiche doganali, ferroviarie; del secondo, i censimenti decennali di molti paesi e i quinquennali di Francia, Germania e Svezia; del terzo, le inchieste politico-amministrative, in quanto contengono speciali rilievi statistici, ad esempio l'inchiesta del 1885 sulle condizioni igieniche e sanitarie dei nostri Comuni. Rilevazioni periodiche, ma separate da intervalli non regolari, sono quelle delle elezioni politiche.

Le rilevazioni continue sono anche automatiche; le altre, riflesse.

d) ***Rilevazione intensiva o monografica*** e ***rilevazione estensiva o statistica in senso stretto***. — Forme precipue della rilevazione intensiva sono le « inchieste » a base statistica e le « monografie » di famiglia, di officina e di Comune, le quali non solo approfondiscono analiticamente l'aspetto formale numerico del fenomeno, ma ne descrivono l'ambiente, la storia, i fattori, gli elementi perturbatori, ecc., entrando in particolarità anche non suscettive di enumerazione. Le ordinarie indagini statistiche hanno invece carattere estensivo.

Il metodo monografico, introdotto dal Le Play per lo studio dell'organizzazione famigliare e dello stato economico delle classi operaie (1), è stato esteso in Francia all'officina (*Monographie d'atelier*) e al Comune (*Monographie de Commune*) (2), ma sembra a taluno sconfinare dal dominio della Statistica, come quello che più non si proporrebbe l'osservazione per masse, ma per tipi. Qui però si pecca d'esagerazione. In realtà non è che si osservino casi assolutamente tipici, che non esistono nell'ordine dei fatti sociali, ma casi quasi-tipici, i quali riuniti, elaborati, confrontati, consentono di generalizzare per induzione, quantunque siano osservati in piccole masse. Siamo dunque ancora in tema di osservazione per masse o che tende a farsi tale man mano che le indagini di questa natura si moltiplicano. Fra un secolo, se il metodo monografico seguiterà a fare proseliti, nessuno più dubiterà del suo carattere statistico.

Il piano o programma delle monografie di famiglia abbraccia le osservazioni sulle condizioni dei singoli membri della famiglia (età, stato civile, produttività o improduttività, qualità e misura del salario, ore di lavoro, modo di alimentazione, ecc.); il bilancio delle entrate e delle spese; infine le note caratteristiche dell'ambiente sociale in cui vive la famiglia (caratteri dell'industria locale, regime di trasmissione dei beni, ecc.). In ragione appunto della grande specialità delle indagini, queste monografie non possono essere che il frutto di private iniziative (3).

Le « inchieste » rientrano nel campo dell'osservazione statistica o ne sconfinano, secondo che i rilievi numerici ne costituiscono il principale fondamento o una illustrazione accessoria. Il loro carattere di rilevazione intensiva, monografica, che si estende alle cause dei fatti in esame, all'ambiente in cui si sono svolti, ecc., non può essere revocato in dubbio. Le inchieste sono private o pubbliche; le pubbliche

---

(1) Veggasi Le Play, *Les ouvriers européens*, Paris 1860-1868, e la Collezione di monografie della « Societé d'économie sociale », di Parigi, dal titolo: *Les ouvriers des deux mondes*.

(2) Veggasi E. Cheysson, *La monographie d'atelier* e *La monographie de Commune* (*Bulletin de l'Institut international de Statistique*, vol. II, fasc. 1° e vol. IX, fasc. 2°).

(3) Veggasi U. Guérin, *De la méthode des monographies de famille* (*Bulletin de l'Institut international de Statistique*, vol. III, fasc. 1°, Roma 1888).

sono governative o parlamentari; queste ultime avendo forza di legge importano l'obbligo di rispondere per gli interrogati.

Un posto intermedio fra le ordinarie statistiche estensive e le intensive o monografiche occupano le investigazioni dette « rappresentative » dal Kiaer, delle quali abbiamo già parlato trattando dei limiti combinati di luoghi e di casi nella rilevazione.

e) *Rilevazione diretta o propria* e *rilevazione indiretta o congetturale*. — È questa una distinzione comunemente adottata dagli statistici, ma fondata a nostro avviso su una confusione di criteri. La rilevazione non può intendersi che nella forma diretta: una enumerazione di casi. Che i casi siano pochi o molti, presi a sorte o scelti con riguardo a caratteri quasi-tipici, rilevati con precisione o per via di stime o misure approssimative, poco importa. Ciò che altri chiama rilevazione indiretta o congetturale è propriamente *induzione* o *deduzione* e deve trovar posto in sede di Statistica, trattata, non come forma di osservazione, ma come forma di ragionamento. Se io calcolo la ricchezza privata esistente in un paese colla sola scorta dei dati delle successioni ereditarie, se dal numero conosciuto delle case argomento quello degli abitanti, ecc., io ragiono, non osservo, non rilevo.

§ **6.** Non potendosi affidare alla memoria dei raccoglitori la massa di casi venuti a loro conoscenza, nè ai loro sensi imperfetti la determinazione delle misure, si ricorre all'impiego di mezzi e strumenti adatti al genere dei rilievi progettati.

Mezzi o modelli di rilevazioni sono i *registri*, *ruoli* e simili per le rilevazioni automatiche; i *bollettini individuali*, le *schede di famiglia*, i *questionari*, ecc. per le rilevazioni riflesse. Nei registri, ruoli, ecc. si prende nota dei fatti nell'ordine di tempo in cui essi vengono portati a conoscenza dell'ufficio raccoglitore. Il bollettino individuale contiene i quesiti concernenti una sola unità statistica; la scheda di famiglia contiene quelli concernenti i diversi membri d'una famiglia. Il questionario è un modello complesso per fatti o situazioni complesse e che si dirama ad Autorità locali, a persone di speciale competenza, ecc. e non proprio ai singoli individui, che di quelle situazioni formano l'intreccio; esso può contenere anche apprezzamenti e giudizi, voti e proposte.

Chiarezza, discrezione, economia delle domande, adattamento delle medesime all'intelligenza media e alla media coltura degli interrogati, alle loro abitudini di sincerità, alle loro diffidenze per tutto ciò che sa d'inchiesta pubblica, tali i requisiti in genere dei modelli di rilevazione. Persino non può dirsi indifferente il dar loro dimensioni più o meno grandi. Per esempio, nei censimenti, ad una scheda di famiglia necessariamente grande per contenere le notizie relative ai diversi componenti, e ingombra di colonne facili a scambiarsi dai poco pratici e dai semianalfabeti, convien preferire d'ordinario il bollettino

individuale, ripetuto per ogni singolo membro della famiglia, ma di interpretazione più sicura.

Strumenti veri e proprii occorrono nelle osservazioni antropometriche (metro, compasso, dinamometro, spirometro, ecc.), nelle osservazioni meteorologiche (barometro, termometro, pluviometro, ecc.) e simili. Non pochi strumenti sono essi stessi dei registratori; l'osservatore non fa che trascrivere ed elaborare i valori indicati automaticamente da quelli. La rivelazione assume qui appunto una forma automatica per eccellenza.

Riguardo agli strumenti si raccomanda l'uniformità di quelli adoperati da diverse persone, fra le quali sia ripartito il lavoro di rilevazione; e mancando l'uniformità bisogna istituire un preventivo calcolo, basato su un certo numero di prove, delle differenze di risultati, cui dànno luogo i diversi strumenti. In certe investigazioni di antropometria sono poi gli stessi oggetti osservati che debbono maneggiare lo strumento misuratore (come il dinamometro per la misura della forza renale o della forza di stringimento della mano, lo spirometro per quella della capacità polmonare, ecc.); ma una sola prova può dare risultati dubbi e convien ripeterla, tenendo conto dei valori massimi raggiunti o dei valori medi, secondo i casi.

Per evitare la disparità di apprezzamenti di diversi osservatori nelle indagini relative al colore degli occhi o dei capelli in una massa di individui (coscritti, scolari, ecc.) giovano le scale cromatiche, che offrono nelle loro gradazioni punti di riferimento ben definiti. Altrimenti i rilievi celano facili errori personali. La stessa gradazione di colore della capigliatura, che pare bionda sulla faccia abbronzata del contadino, può apparire castana sul viso bianco-roseo di un borghese.

§ 7. A titolo di semplice illustrazione pratica delle cose dette, indichiamo il metodo di rilevazione seguito in alcune statistiche italiane, accompagnandovi qualche modello di scheda o di questionario.

a) Movimento della popolazione secondo gli atti dello stato civile. — Le notizie dei matrimoni e delle morti sono fornite dagli uffici comunali di stato civile sopra schede individuali e lo spoglio di queste si fa direttamente a cura della Direzione generale della Statistica in Roma. Quanto alle nascite, i Comuni inviano alla Direzione generale anzidetta un prospetto riassuntivo del numero dei nati e dei natimorti in ciascun mese dell'anno, divisi per sesso e secondo l'origine legittima o illegittima, come pure del numero dei figli naturali legittimati e dei parti multipli.

(MODELLO DI SCHEDA PER MATRIMONI).

Notizie fornite dall'ufficiale di stato civile.

Marito.	*Moglie.*
Cognome e nome..... Professione o condizione..... Stato civile (celibe o vedovo)..... Età (anni compiuti)..... Se sottoscrisse l'atto (rispondere per *sì* o per *no*).....	Cognome e nome..... Professione o condizione..... Stato civile (nubile o vedova)..... Se è vedova, dicasi se ha figli minorenni..... Età (anni compiuti)..... Se sottoscrisse l'atto (rispondere per *sì* o per *no*).....

In caso di matrimonio fra *consanguinei*, indicare il grado di parentela (*tra zio e nipote, tra zia e nipote, tra cugini in primo grado*).

A tergo della scheda si hanno poi le indicazioni del numero d'ordine del registro, del Comune e circondario, dell'anno e mese e infine la firma dell'ufficiale di stato civile.

b) *Statistica delle cause di morte.* — Si fa raccogliendo sopra una scheda nominativa per ciascun defunto la dichiarazione della malattia che causò la morte, dichiarazione compilata e firmata dal medico curante o, in mancanza di esso, dal medico necroscopo, che ha dato il permesso di seppellimento; per i bambini morti poco dopo il parto le dichiarazioni si fanno dalle levatrici.

(MODELLO DI SCHEDA PER LA STATISTICA DELLA MORTALITÀ).

Maschi.

Certificato medico di morte.	*Notizie fornite dall'ufficiale di stato civile.*
Cognome e nome del defunto..... Domiciliato in via..... N°..... *Dichiaro che la causa della morte del sopranominato, secondo la mia scienza e coscienza, fu quella sottoindicata:* Morte naturale { Malattia prima..... / Accidente terminale..... Morte violenta { Accidentale..... / Suicidio..... / Omicidio..... Firma e qualifica	1. Età, anni..... Pei bambini che avevano meno di un anno: Mesi..... giorni..... 2. Se il bambino aveva meno di 5 anni dicasi se era legittimo..... illegittimo..... 3. Stato civile..... 4. Se il defunto era vedovo o aveva figli legittimati o riconosciuti, dicasi se ha lasciato figli minorenni..... 5. Professione o condizione..... (per gli adulti da 15 anni in su). 6. Aveva dimora stabile nel comune? ..... od occasionale?.....

(Analoga scheda, in caratteri rossi, per le femmine).

Le dichiarazioni originali si spediscono ogni mese dai sindaci dei singoli Comuni, pel tramite della Prefettura, all'Ufficio centrale di Statistica (Direzione generale), dove vengono esaminate da un medico, il quale contrassegna ciascuna di esse con un numero corrispondente alla voce analoga di una classificazione prestabilita delle cause di morte.

c) *Statistica dell'emigrazione.* — Fonte principale i registri dei passaporti e, sussidiariamente, la notorietà. I sindaci e le Autorità di pubblica sicurezza, per poter distinguere gli emigranti dai viaggiatori propriamente detti, prima di rilasciare il passaporto s'informano se i richiedenti si recano all'estero per cercarvi lavoro ovvero per diporto, per affari, per ragion di studio, ecc. Questi ultimi, a differenza dei primi, appartengono generalmente alle classi agiate e pagano una tassa per ottenere il passaporto, laddove questo viene rilasciato gratuitamente agli altri, che costituiscono la quasi totalità dei nostri emigranti. Oltre a ciò le Autorità politiche locali procurano d'aver notizie delle persone che si allontanano dal paese per andare all'estero senza regolare passaporto, perchè renitenti alla leva o ricercati dall'Autorità giudiziaria, ecc.

d) *Statistica delle elezioni generali politiche.* — I dati relativi alle elezioni generali politiche del 1900 vennero forniti per la massima parte dalle Cancellerie dei tribunali, che li ricavarono dalle copie dei verbali depositate presso di esse. Dei pochi collegi, pei quali non furono depositate le copie dei verbali o queste riuscirono incomplete, le notizie si richiesero ai prefetti. I dati ricavati dai verbali sono da considerarsi come definitivi, salve le rettificazioni che può portarvi la Camera nella verifica dei poteri; quelli forniti dai prefetti sono approssimativi.

e) *Statistica giudiziaria penale.* — Per questa i dati vengono raccolti in parte con *registri numerici* riempiti giorno per giorno negli uffici delle Cancellerie e Segreterie giudiziarie e in parte col mezzo di *schede descrittive* compilate negli stessi uffici a processo esaurito. La scheda serve a far conoscere il numero e la specie dei reati, l'esito definitivo delle istruttorie e dei giudizi e inoltre le qualità personali, i precedenti penali e la recidività degli imputati. I registri giornalieri sono intesi ad accertare il numero delle ordinanze e sentenze pronunziate, la durata dei procedimenti, la durata della carcerazione preventiva, i provvedimenti sulla libertà provvisoria e simili.

f) *Antropometria militare.* — L'Ispettorato di sanità militare al Ministero della guerra in Roma ha eseguito recentemente lo spoglio dei fogli sanitarii degli arruolati delle classi 1859-1863 ed ha pubblicato un primo volume di dati antropologici ed etnologici, cui un altro seguirà di dati igienici e medico-legali. Il foglio sanitario, adottato su proposta del dott. Guida, doveva servire a raccogliere i dati soliti a prendersi alla prima visita dell'arruolato e inoltre le annotazioni di

tutte le vicende sanitarie del medesimo durante la sua carriera militare. Noi riportiamo qui solo la pag. 3 di detto foglio (che ne comprendeva 8), come quella che concerne le qualità fisiche invariabili o poco variabili, le quali formarono la principale materia di spoglio pel primo volume (1).

QUALITÀ FISICHE INVARIABILI O POCO VARIABILI.

Capelli { colore	Sopracciglia .....
Capelli { forma	Fronte .....
Occhi	Naso .....
Colorito	Bocca .....
Dentatura	Mento .....
Segni particolari	Viso .....

Diametri del capo { antero-posteriore massimo, centim. .....
Diametri del capo { trasversale » » .....

QUALITÀ FISICHE VARIABILI	ANNI DI SERVIZIO					
	1°	2°	3°	4°	5°	Congedo
Statura . . . . metri	...	...	...	...	...	...
Perimetro toracico . . . . »	...	...	...	...	...	...
Peso . . . . chilogr.	...	...	...	...	...	...
Firma dell'ufficiale medico . .	...	...	...	...	...	...

Difetti fisici non esimenti dal servizio all'atto dell'arruolamento.....
Variazioni avvenute in essi.....

Vaccinazione. — Stato antecedente (*).....

VACCINAZIONE E RIVACCINAZIONE	VACCINO		Data	Esito *	FIRMA dell'ufficiale medico
	animale	umanizzato			
Primo innesto	...	...	...	...	.....
Secondo »	...	...	...	...	.....

(*) Vaiolato — Vaccinato — Non vaccinato.

g) ***Statistica degli scioperi.*** — Le notizie sugli *scioperi* e sulle *chiusure* (2) avvenuti nel 1900 furono raccolte dai prefetti, ai quali si richiesero mediante un questionario. Le 15 domande del questionario,

(1) V. *Antropometria militare*, parte 1ª, *Dati antropologici ed etnologici*, pag. 10-14, Roma, presso il *Giorn. medico del R. Esercito*, 1896. La direzione dei lavori di questa importantissima inchiesta fu affidata al dott. Ridolfo Livi.

(2) «Chiusura» è la sospensione temporanea di lavoro deliberata dal proprietario o imprenditore di propria iniziativa, per costringere gli operai ad accettare le condizioni da lui imposte o per solidarietà che egli intenda

che per brevità riportiamo senza gli sviluppi e i chiarimenti annessi, erano le seguenti:

1° Industria nella quale avvenne lo sciopero.

2° Mese e giorno in cui cominciò e mese e giorno in cui finì lo sciopero.

3° Occupazione o condizione degli operai che iniziarono lo sciopero.

4° Denominazione delle officine o stabilimenti ovvero della proprietà in cui si verificò lo sciopero o la chiusura; numero degli operai impiegati prima dello sciopero e numero di quelli che scioperarono, distinti in adulti e fanciulli, maschi e femmine.

5° Numero degli operai, distinti come sopra, che furono costretti a sospendere il lavoro, non di propria iniziativa, ma in conseguenza dello sciopero di altri operai (*perchè, ad esempio, venne a mancare la forza motrice o la materia prima*) ovvero per altra cagione indipendente dalla loro volontà.

6° Cause dello sciopero degli operai (domanda di aumento di salario o di diminuzione d'orario, opposizione ad una diminuzione di salario o ad un aumento di ore di lavoro, ecc.).

7° Cause della chiusura operata dal proprietario.

8° e 9° Vicende ed esito dello sciopero o della chiusura.

10° Mercedi giornaliere dei lavoratori *prima* dello sciopero o della chiusura e *dopo*.

11° Orario del lavoro prima e dopo lo sciopero o la chiusura.

12°-15° Se esistevano leghe di resistenza, se gli scioperanti ricevettero sussidi durante lo sciopero; se vi fu turbamento dell'ordine pubblico; se precedenti scioperi si verificarono nello stesso stabilimento o nella stessa proprietà.

h) *Inchiesta sulle condizioni del lavoro nelle risaie della Lomellina, del Vercellese e del Novarese* (1). — Di questa interessante inchiesta,

---

di mantenere coi padroni di altri stabilimenti, ovvero per punire gli operai di qualche infrazione ai regolamenti o infine per protestare contro provvedimenti fiscali.

Osserva la nostra Direzione generale di Statistica che non è sempre facile distinguere lo *sciopero volontario* fatto dagli operai dalla *chiusura*, perchè spesso la sospensione del lavoro partecipa dell'uno e dell'altra, o almeno cominciata per volontà degli operai si trasforma poscia in chiusura per volontà dei proprietari. Il criterio giustamente seguito dalla Direzione generale della Statistica è quello di classificare come sciopero la sospensione di lavoro, quando la prima iniziativa è partita dagli operai, senza preoccuparsi se nelle fasi successive le parti siano state invertite, cioè se il proprietario abbia ordinato la chiusura anche a danno di quei reparti di operai che non avevano da principio partecipato allo sciopero.

(1) Dott. G. LORENZONI, *I lavoratori delle risaie* (Pubblicazione dell'Ufficio del lavoro della Società Umanitaria), Milano 1904. Parte I, Introduzione generale e Relazione riassuntiva dell'inchiesta sulla mondatura.

eseguita a cura della Società Umanitaria di Milano per mezzo di appositi impiegati rilevatori, diamo, ancora in forma abbreviata, il questionario preparato dal professore G. Montemartini e concernente la *monda del riso.*

Notizie generali. (Provincia, Comune, Frazione — Denominazione della cascina — Nome e cognome del proprietario o castaldo — Se vige qualche regolamento provinciale per la coltivazione del riso).

La Cascina. (A quale distanza trovasi dal perimetro della risaia; se vi hanno accumulazioni di materie d'escavazione o di concime, se l'acqua dei pozzi è sana, ecc.).

La Risaia. (Superficie coltivata a risaia stabile o a vicenda; decorso libero o non libero delle acque nei fossi di scolo, uso di concimi chimici, ecc.).

Composizione della squadra. (Se di lavoratori locali o di immigrati o mista; numero dei lavoratori distinti per sesso e in tre classi d'età; provenienza degli immigrati; quanti dei lavoratori occupati sono inscritti in leghe di resistenza od unioni professionali).

Condizioni effettive del lavoro. 1° L'*orario* (inizio e fine, riposi, ecc.).

2° Il *salario* (salario giornaliero delle varie categorie, in denaro e in natura; termini di pagamento, ecc.; durata del lavoro di monda e guadagno complessivo di un uomo, di una donna e di un fanciullo; lavori supplementari e modo di retribuzione).

3° Il *nutrimento* (se si provvede al nutrimento della squadra individualmente o collettivamente; dettagli sulla composizione dei pasti; da chi son forniti i generi alimentari; se il nutrimento è sano; in caso di cibi malsani raccogliere possibilmente campioni; se l'acqua è sana, in quali giorni si beve vino e quanto per individuo).

4° I *dormitorii* (come dorme ed è alloggiata la squadra; se in luoghi chiusi, dire delle condizioni di pavimentazione, aereazione, ecc.; se in luoghi aperti (portici, fienili); se in stalle o rimesse; condizioni igieniche di questi locali; se dormitorii separati per uomini e per donne).

5°-8° *Profondità dell'immersione nell'acqua durante i lavori; se si usano calze protettrici; conseguenze del non uso; casi d'insolazione e di febbri malariche. Cenni generici sulla influenza del lavoro di risaia sulla salute e moralità dei lavoratori.*

Contratti di lavoro. A) *Contratto diretto fra il conduttore del fondo e la squadra* (epoca di stipulazione, condizioni di pagamento, durata, forma collettiva o individuale, variazione di patti in confronto di anni precedenti, distinguendo fra squadre indigene e immigrate).

B) *Contratto indiretto per interposizione dell'incettatore* (rapporti fra conduttore e incettatore o caporale; patti che legano quest'ultimo

al padrone, garanzie reciproche; profitto dell'incetta, rapporti fra incettatore o caporale e lavoratori; tempo e modo di arruolamento; garanzie contrattuali; a carico di chi stanno le spese di trasporto; se furono osservati i patti relativi al vitto, all'orario e all'alloggio; se ve ne sono di relativi al rimpatrio per malattia, ecc.

Leggi protettrici del lavoro. (Se e in quali casi non furono osservate le leggi sanitarie e i regolamenti dei lavori di risaia e la legge sul chinino di Stato; provvedimenti adottati dalle Autorità governative, provinciali e comunali, dalle Leghe e Federazioni di resistenza, dalle Unioni professionali).

Non possiamo chiudere queste brevi e pratiche illustrazioni di metodi di rilevazione, senza accennare all'opera di quel nuovo nostro osservatorio economico, che è l'Ufficio del lavoro presso il Ministero d'agricoltura, industria e commercio. Le informazioni sul *mercato del lavoro* sono fornite al detto Ufficio dalle Camere di commercio e da quelle del lavoro, da Associazioni di industriali, da Federazioni di mestiere, da Cattedre ambulanti e Scuole d'agricoltura, da prefetti, sindaci e agenti consolari. In brevi, ma assai ragguardevoli « Note metodologiche » il prof. G. Montemartini, direttore dell'Ufficio, ha tracciato il programma delle indagini, che qui riassumiamo (1).

Il mercato del lavoro suppone la rilevazione dei seguenti elementi: salari, domanda e offerta di lavoro, orari, organizzazione delle forze contraenti, migrazioni, conflitti di lavoro. La determinazione della domanda e dell'offerta di lavoro si concreta in quella del numero degli occupati e dei disoccupati.

Soffermandoci qui solo al fenomeno della disoccupazione, due metodi si presentavano all'ufficio: il metodo, che possiamo dire tedesco, il quale studia l'*offerta di lavoro* e cerca di determinare il numero dei disoccupati dallo stato di questa offerta; e il metodo inglese, che invece parte dalla *domanda*. Il primo consiste tutto nel registrare per ogni piazza e per ogni industria il numero dei disoccupati che si presentano agli uffici di collocamento. Accanto al numero dei posti richiesti dai lavoratori si registra quello dei posti offerti dagli industriali; la differenza tra le due cifre segna la disoccupazione. Per l'efficacia del metodo occorre dunque un sistema di *mediazione del lavoro* quasi perfetto, quale appunto si ha in Germania. Il metodo inglese invece, basato sullo studio della domanda, esige che si determinino le contrazioni, le espansioni, il ritmo delle diverse industrie. Data la conoscenza dei diversi opifici di una regione, si cerca agli imprenditori, mese per mese, qual è il numero dei lavoranti da essi

(1) V. *Bollettino dell'Ufficio del lavoro*, vol. I, n. 1 e 2, aprile-maggio 1904 pag. 40 e seguenti. Ministero di agricoltura, industria e commercio, Roma, Bertero, 1904.

impiegato. Confrontando così i diversi periodi con un periodo-base, si ottiene il numero dei disoccupati ad ogni contrazione dell'industria. Questo metodo è attuabile quando si abbia a fare con industriali intelligenti e volenterosi o meglio ancora con forti Associazioni di industriali, come è precisamente il caso dell'Inghilterra.

La risoluzione adottata dal nostro Ufficio del lavoro è eclettica. Esso cerca di determinare il fenomeno, seguendo i due sistemi descritti, e contemporaneamente di sviluppare gli organismi che dànno vita all'uno e all'altro sistema, utilizzando gli elementi offerti dai pochi Uffici di collocamento che funzionano in Italia e quelli offerti dalle organizzazioni padronali e di lavoratori. Dei risultati di questo procedimento è prematuro parlare; ma tutto lascia credere ch'essi corrisponderanno all'aspettazione degli studiosi di questioni sociali.

TITOLO II.

LA FORMAZIONE DELLE TABELLE PRIMITIVE.

Sommario: § 1. Numerazione, spoglio e aggruppamento; tabelle provvisorie. — § 2 e 3. Tabelle definitive, semplici e complesse. — § 4. Serie e seriazioni.

§ **1.** Alla raccolta seguono la numerazione dei modelli originari della rilevazione, lo spoglio e l'aggruppamento dei dati. La ***numerazione*** ha per iscopo di stabilire l'eguaglianza o meno dei modelli diramati e di quelli pervenuti, massime nel caso di risposte volontarie; inoltre di fissare il termine più generale di riscontro per l'esattezza quantitativa delle diverse classificazioni che si faranno in seguito, il *tutto* dovendo essere eguale alla somma delle sue *parti*, comunque venga frazionato. La numerazione si fa espressamente per i bollettini individuali, le liste o schede di famiglia, i questionari; non è necessaria per i registri che già contengono un numero d'ordine a meno che le loro caselle siano riempite in maniera saltuaria, come accade nella registrazione dei votanti man mano che si presentano alle rispettive sezioni nelle elezioni politiche o amministrative.

Lo *spoglio* è l'operazione per la quale le singole unità statistiche si trasportano dai modelli originari, che non permettono per se stessi l'aggruppamento ad altri modelli che lo permettono. Questi secondi modelli sono *tabelle provvisorie* divise in tante colonne quante le specie volta per volta rilevate del fenomeno, ogni caso o caratteristica del quale viene, come unità statistica, segnato nella competente colonna con un punto o una lineetta o un numero nell'ordine naturale di successione dei numeri interi. Questo trasporto si opera in maniera immediata per i bollettini individuali e anche per le schede di famiglia, finchè la « famiglia » vien trattata come unità; invece pei registri e

per le stesse schede di famiglia, quando della famiglia vogliam considerare i singoli individui, conviene riprodurre prima i dati individuali su cartoline individuali e poi operare il trasporto degli elementi caratteristici sulle tabelle provvisorie nella maniera già indicata. La qual maniera è suscettiva di una notevole semplificazione. Invece di annotare unità per unità, si può con risparmio di tempo formare tanti separati gruppi e sottogruppi di modelli originari, quante le classi e sottoclassi prestabilite, contare le unità d'ogni gruppo e riportare la somma sulla tabella all'uopo predisposta. In qualche circostanza i due metodi, quello semplificato e quello non semplificato, si combinano. Così in molti uffici elettorali, nelle elezioni amministrative che si fanno a scrutinio di lista, si registrano unità per unità le sole schede manoscritte e quelle stampate che contengono qualche cancellazione o sostituzione di nomi; mentre le stampate, che a scorsa d'occhio si riconoscono uniformi, vengono riunite in gruppi omogenei e poi contate e registrate in blocco.

Lo spoglio e l'aggruppamento secondo il piano prestabilito di classificazione procedono di conserva. Talora però qualche classificazione non preveduta è suggerita dallo stesso spoglio; tal'altra una classificazione progettata si rivela ai primi assaggi poco concludente o troppo laboriosa e perciò viene sacrificata.

Di regola, spoglio e aggruppamento si fanno per elementi *combinati* e non per elementi *isolati*. Ragioni teoriche e pratiche lo consigliano. Scientificamente i dati che risultano dalla prima maniera hanno valore investigativo, in quanto pongono in evidenza delle correlazioni, dei punti di massimo e di minimo, delle forme di distribuzione caratteristiche; quelli, invece, che si ottengono da uno spoglio per elementi isolati hanno appena valore descrittivo. Se ad esempio, io ho enumerato per tutto un paese gli individui nati nello stesso Comune in cui risiedono, i nati in un Comune diverso da quello in cui risiedono, ma appartenente allo stesso circondario o almeno alla stessa provincia, o almeno alla stessa regione, ecc., e poi, *senza più tener distinte le località di nascita*, conto quanti sono gli individui addetti all'agricoltura, all'industria manifattrice, al commercio, alle professioni ed arti dette liberali, alla milizia, ai servizi pubblici civili, avrò due specie di dati in certo modo indipendenti e di valore appena descrittivo, inetti per se medesimi a farmi conoscere la correlazione importantissima che intercede fra il genere di professione esercitata e la maggiore o minore sedentarietà, la maggiore o minore migrabilità e mescolanza dei gruppi di popolazione. Operando, invece, lo spoglio per elementi combinati, cioè distinguendo ad un tempo i censiti per professione e per località di nascita comparata a quella della residenza, mi assicuro un ordine ben più esteso di cognizioni in materia. Le ragioni pratiche sono poi queste; il lavoro di spoglio non cresce, siccome a prima giunta sembrerebbe, come il prodotto delle classi o categorie che si combinano, ma cresce assai meno rapidamente, essendo

il rilievo quasi simultaneo per i due, tre o quattro elementi da combinarsi, purchè, beninteso, non si vada oltre i tre o i quattro, sotto pena di rendere difficile e lenta la ricerca della colonna in cui va fatta l'annotazione, e probabili gli errori o scambi di posto.

Facciamo un esempio di combinazione di tre caratteri, come l'età, il sesso e lo stato civile. Lo spoglio non può riuscire difficile, perchè limitato è il numero delle categorie cui dànno luogo gli elementi *sesso* e *stato civile*, e quanto all'*età* si possono prendere intervalli abbastanza larghi, come di cinque in cinque anni. Il casellario non apparirà certo ingombrante:

CLASSI DI ETÀ	MASCHI				FEMMINE				Totale generale
	Celibi	Coniugati	Vedovi	Totale	Nubili	Coniug.	Vedove	Totale	
Da 0 a 5 anni	...	...	...	...	...	...	...	...	...
Da 5 a 10 »	...	...	...	...	...	...	...	...	...
Da 10 a 15 »	...	...	...	...	...	...	...	...	...
Da 15 a 20 »	...	...	...	...	...	...	...	...	...
ecc., ecc.									
Totali	...	...	...	(A)	...	...	...	(B)	(C)

Il totale registrato nella casella (A) deve risultare identico tanto dall'addizione dei parziali contenuti nelle caselle superiormente disposte, quanto da quella dei parziali contenuti nelle tre caselle orizzontali alla sua sinistra. Così dicasi del totale registrato nella casella (B). Il totale generale scritto nella casella (C) deve risultare identico così dalla somma di (A) e di (B), come dalla somma dei numeri contenuti nelle caselle disposte superiormente ad esso. Si ha così un mezzo di riscontro dell'esattezza delle operazioni aritmetiche.

Norme pratiche da seguire sono ancora queste: raccolte e debitamente numerate le schede, si adotti come criterio per la prima loro separazione in gruppi l'elemento che dovrà ricomparire il maggior numero di volte nelle combinazioni progettate e si tenga ferma quella separazione fino ad esaurimento di dette combinazioni. Inoltre, delle combinazioni di elementi due a due, si tengano ferme quelle che ricompariranno maggior numero di volte nei gruppi di elementi presi tre a tre; il che implica che si ricompongano o si rifondano i gruppi, dei quali un elemento non dovrà più ricomparire nelle successive combinazioni; e così di seguito. Beninteso che le esigenze speciali di certi rilievi statistici possono consigliare altri spedienti.

La graduazione regolare delle categorie, quando gli elementi, che le determinano, sono suscettibili di espressione quantitativa, semplifica molto il lavoro di spoglio. È nota la comodità che offrono le cifre rotonde. Così in una classificazione per età si potrà, poniamo, stabilire una scala in questo modo: da 0 a 5 anni, da 5 a 10, da 10 a 15, ecc., non già prendendo per limite numeri non interi di anni, nè facendo ad arbitrio gruppi ora quinquennali, ora triennali, decennali, ecc. Il che non vuol dire che si debba tutto sacrificare all'uniformità; anzi potrà convenire di specificare maggiormente le categorie di maggior importanza o quelle nelle quali il fenomeno presenta condizioni più caratteristiche di variabilità. Nella statistica della mortalità la classificazione per età si può far procedere così: dalla nascita a 1 mese, da 1 a 3 mesi, da 3 a 6, da 6 a 9, da 9 a 12; poi si continua d'anno in anno fino a 5; poi ancora di cinque in cinque anni e dopo un certo punto anche di dieci in dieci anni.

Non approvabile in certi casi è l'uso di formare categorie senza limiti estremi indicati. Per esempio, in una classificazione di depositi a risparmio presso una Banca si dice: tanti depositi di ammontare inferiore a 100 lire, tanti da 100 a 500, da 500 a 1000, da 1000 a 2000, fino ad un'ultima categoria, la quale comprende, poniamo, i depositi da 5000 lire in su. Sicchè non si conosce nè il deposito di minimo, nè quello di massimo importo, che è quanto dire il punto iniziale e quello terminale della seriazione, la cui rappresentazione grafica non si potrebbe fare, come vedremo, senza ricorrere ad ipotesi ed artefici. Le categorie indeterminate nascondono inopportunamente i fenomeni rari (matrimoni tra giovanissimi e vecchissimi, suicidii di ragazzi, stature nane e giganti, ecc.), che pur sono materia di interessanti applicazioni del calcolo delle probabilità.

Lo spoglio dev'essere eseguito in maniera da permettere non solo la discriminazione dei casi estremi, dei fenomeni rari e caratteristici, ma anche, se possibile (quando si tratti di fenomeni assai variabili nello spazio), in maniera da permettere la presentazione del fatto collettivo sotto altro aspetto che non sia quello della semplice ripartizione amministrativa per provincie e circondari. Il metodo cosiddetto « geografico », raccomandato dal Mayr (1), consiste appunto nel presentare la distribuzione territoriale del fatto secondo le caratteristiche naturali o sociali dei diversi ambienti; quindi nel contrasto fra la città e la campagna, fra il monte e il piano, le coste e l'interno, le plaghe asciutte e quelle visitate da frequenti pioggie, quelle di clima caldo e quelle di clima freddo, ecc. La qual cosa appare ancora fattibile quando i dati sono forniti per divisioni amministrative, se queste discendono fino al Comune o almeno al mandamento. Per le grandi città con-

(1) G. von Mayr, *Statistik und Gesellschaftslehre*, vol. I, *Theoretische Statistik*, pag. 86 e 87, Freiburg 1895.

viene che lo spoglio tenga distinti i quartieri in modo da permettere i successivi raggruppamenti dei dati secondo la zona interna, la periferica e l'intermedia o secondo la divisione sociale di quartieri: ricchissimi, ricchi, medi, poveri, poverissimi. Ma su questo tema ci converrà ritornare parlando della comparazione dei dati.

§ 2. Le *tabelle definitive* sono prospetti o quadri nei quali i risultati dello spoglio vengono disposti in tante colonne verticali quante sono le varietà del fenomeno osservato, e in tante colonne orizzontali quanti i momenti o luoghi o condizioni in cui esse varietà furono rilevate. Le tabelle definitive non differiscono dalle provvisorie, se non perchè riportano le somme già fatte delle unità statistiche. Il quadro seguente espone, a titolo di esempio, la natalità in Italia considerata nelle sue specie (elementi combinati: il sesso e la filiazione legittima o illegittima) e nella successione del tempo (gli anni dal 1897 al 1902):

Anni	LEGITTIMI		ILLEGITTIMI riconosciuti		ILLEGITTIMI non riconosciuti ed esposti		TOTALE
	Maschi	Femmine	Maschi	Femmine	Maschi	Femmine	
1897	530.891	500.758	21.363	20.003	14.400	14.433	1.101.848
1898	515.417	487.395	20.238	18.662	14.275	14.087	1.070.074
1899	524.133	497.573	20.517	18.897	13.781	13.657	1.088.558
1900	515.360	488.610	19.230	17.790	13.346	13.040	1.067.376
1901	512.452	484.023	18.428	17.037	13.127	12.696	1.057.763
1902	529.400	501.143	19.179	17.215	13.091	13.046	1.093.074

Se invece di una successione di tempi avessimo da considerare il fenomeno nella sua distribuzione geografica, in un anno dato, l'ordine degli Stati o quello dei compartimenti, provincie o Comuni di un medesimo Stato si disporrebbe ancora nella prima colonna a sinistra. Per esempio, nel 1901 la natalità legittima e illegittima per compartimenti in Italia fu la seguente:

COMPARTIMENTI	LEGITTIMI		ILLEGITTIMI riconosciuti		ILLEGITTIMI non riconosciuti ed esposti		TOTALE
	Maschi	Femmine	Maschi	Femmine	Maschi	Femmine	
Piemonte	48.804	46.250	724	664	761	700	97.903
Liguria	14.259	13.757	409	416	501	412	29.754
Lombardia . . .	75.599	71.494	773	753	965	995	150.579
Veneto	56.179	53.006	2.251	2.063	1.011	880	115.390
Emilia	36.863	34.995	3.688	3.413	1.267	1.205	81.431
ecc.							

All'ordine alfabetico degli Stati, dei compartimenti, ecc., è da preferire in generale la successione geografica, poniamo, da nord a sud, ovvero la successione con riguardo alla maggiore o minore intensità colla quale si manifesta il fenomeno. Questo secondo modo però si applica non alle tabelle primitive, ma alle derivate.

Si potrebbe nel primo dei precedenti prospetti porre la successione dei tempi (e nel secondo la successione dei compartimenti) in testa alle colonne verticali, e invece le varietà del fenomeno nella prima colonna a sinistra. Tuttavia, a parità d'altre circostanze, riesce più comoda all'occhio la disposizione sopra adottata, perchè i numeri costituenti una serie continua cronologica o geografica occupano una linea più breve quando siano collocati l'uno sotto l'altro, che non quando lo siano uno al fianco dell'altro; ed è appunto la continuità della serie che più importa abbracciare con uno sguardo. Così delle due serie seguenti, verticale l'una, orizzontale l'altra:

Anni	Nati illegittimi in Italia
1897	70.199
1898	67.262
1899	66.852
1900	63.406
1901	61.288
1902	62.531

Nati illegittimi in Italia.

1897	1898	1899	1900	1901	1902
70.199	67.262	66.852	63.406	61.288	62.531

la prima è meglio e più presto compresa e percorsa dall' occhio; il che dipende forse dall'abitudine contratta alla lunga colle operazioni aritmetiche, in cui i numeri hanno generalmente una disposizione verticale. La disposizione orizzontale, che obbliga l'occhio a una certa lentezza e a procedere in certo modo per salti, conviene invece per le classi o varietà in cui è distinto il fenomeno in esame, massime quando abbiano la caratteristica della discontinuità. Del resto la regola patisce eccezioni, subito che il numero delle classi o varietà ecceda notevolmente quello dei luoghi o tempi.

§ **3.** Siccome sul piano non disponiamo che di due dimensioni, lunghezza e larghezza, così riesce impossibile figurare il fenomeno, con una tabella *semplice*, in funzione di più di due elementi. Lo si può, ma con qualche scapito della chiarezza, con una tabella *complessa*. I due prospetti precedenti offrono già un primo grado di complessità, grazie alle tre variabili in essi considerate; non si può passare dai maschi legittimi ai maschi illegittimi riconosciuti e ai non riconosciuti od esposti, se non saltando le colonne assegnate alle femmine. Proviamoci a introdurre un ulteriore grado di complessità. Vogliasi, ad esempio, considerare ancora la natalità italiana dal quadruplice punto di vista del sesso, della origine legittima o illegittima, della distribuzione geografica (per compartimenti) e del tempo, cioè per i diversi anni di un periodo. Bisognerebbe dare questa forma al quadro.

NATI VIVI IN ITALIA NEL TRIENNIO 1899-1901.

COMPARTIMENTI	Anni	LEGITTIMI		ILLEGITTIMI riconosciuti		ILLEGITTIMI non riconosciuti ed esposti		TOTALE
		Maschi	Femmine	Maschi	Femmine	Maschi	Femmine	
Piemonte . . .	1899	47.952	45.843	760	684	765	745	96.749
	1900	48.202	45.284	706	685	796	787	96.460
	1901	48.804	46.250	724	664	761	700	97.903
Liguria.	1899	14.002	13.431	393	363	484	423	29.096
	1900	14.188	13.521	374	366	423	434	29.306
	1901	14.259	13.757	409	416	501	412	29.754
Lombardia . .	1899	72.468	68.626	798	834	1.022	956	144.704
	1900	72.639	69.052	776	747	981	1.005	145.200
	1901	75.599	71.494	773	753	965	995	150.579
ecc.								

La tabella non è semplice; si direbbe piuttosto una fusione di tante tabelle semplici, quanti sono gli anni considerati e quante le sottospecie (i sessi). La continuità dei dati vi appare interrotta; le cifre della natalità riguardanti il Piemonte nel 1901 sono susseguite nell'ordine verticale da quelle della Liguria pel 1899. Per collegare i dati dei varii compartimenti per un medesimo anno l'occhio dovrebbe scorrere il prospetto, direm così, a salti. Perciò questa forma complessa di intavolazione dei dati è ammissibile solo nei casi in cui l'economia di spazio si fa desiderare anche con qualche scapito della chiarezza e comodità della lettura.

Le tabelle, delle quali si è parlato sin qui, si chiamano *primitive*, come quelle che contengono in forma di numeri assoluti i risultati greggi delle operazioni di spoglio e aggruppamento. Hanno invece nome di *derivate* quelle che espongono dati elaborati con procedimenti speciali: medie, rapporti e percentuali, numeri perequati e interpolati, numeri-indici. In un medesimo prospetto possono trovarsi accompagnati i dati primitivi e i derivati; avremo così tabelle *miste*. Le derivate presentano difficoltà particolari di costruzione, che esamineremo più tardi.

§ 4. Gia sappiamo che cosa s'intenda per *unità statistica*. Un complesso di unità per certi rispetti omogenee e formanti quindi una classe, viene sotto il nome di *quantità* o *dato statistico*. Più quantità statistiche concernenti lo stesso fenomeno e riferite ad una successione di tempi costituiscono una *serie*. Più quantità statistiche esprimenti varietà graduate di una stessa specie e quindi riferite ad una determinata scala

di misura, costituiscono una *seriazione*. Se la classificazione procede non per varietà graduabili, ma per elementi discontinui, si ha un *quadro* o *prospetto* statistico in senso stretto.

Delle « serie » altre sono *periodiche* e altre *non periodiche*. Periodiche quelle delle quali si considera lo svolgimento in un ciclo naturale o convenzionale di tempo (come i dodici mesi dell'anno, i sette giorni della settimana, le ventiquattro ore del giorno) e i cui massimi e minimi si ripetono nei successivi cicli ad intervalli abbastanza regolari. Non periodiche quelle nelle quali si prescinde dal ciclo e si considera la semplice continuità di manifestazione del fatto nel succedersi del tempo. Le une e le altre possono avere carattere statico o carattere dinamico; *statico*, se il fenomeno, astrazion fatta dalle piccole oscillazioni che si ha ragion di credere prodotte da cause minute e accidentali o dalle oscillazioni caratteristiche per la loro periodicità, non accenna in definitiva a crescere o a diminuire in funzione del tempo; *dinamico*, in caso diverso. I procedimenti d'interpolazione delle serie mirano poi a specificare le forme della dinamica del fenomeno; movimento rettilineo ascendente o discendente, parabolico di 2° grado, di 3° grado, ecc., come apprenderemo più oltre.

Diamo intanto esempi di alcune serie tratte da statistiche italiane ed estere:

SERIE NON PERIODICHE

Quinquenni	Produzione mondiale dell'oro (medie annuali in Kg.)	Commercio dell'Italia coll'estero			Scioperi in Italia		Esercizi finanziari	Consumo del tabacco da fiuto in Italia (in quintali)
		Anni	Importazione (milioni di lire)	Esportazione (milioni di lire)	nelle industrie propr. dette	nella agricoltura		
1851-1855	199.388	1893	1.191	964	131	18	1892-1893	29.183
1856-1860	201.750	1894	1.095	1.027	109	8	1893-1894	28.775
1861-1865	185.057	1895	1.187	1.038	126	7	1894-1895	27.800
1866-1870	195.027	1896	1.180	1.052	210	1	1895-1896	26.863
1871-1875	173.904	1897	1.192	1.092	217	12	1896-1897	26.355
1876-1880	166.095	1898	1.413	1.204	256	36	1897-1898	25.813
1881-1885	153.643	1899	1.507	1.431	259	9	1898-1899	25.759
1886-1890	169.862	1900	1.700	1.338	383	27	1899-1900	25.373
1891-1895	245.181	1901	1.718	1.374	1.042	629	1900-1901	24.899
1896-1900	*395.000*	1902	1.776	1.472	780	228	1901-1902	24.408
—	—	1903	1.862	1.517	528	45	1902-1903	23.454

Le « seriazioni », dicevamo, suppongono il riferimento dei casi di un fenomeno ad una scala graduata di misura. Così se distribuisco i proprietari di immobili in un paese secondo l'estensione dei terreni, i contribuenti secondo l'ammontare del reddito tassato o dell'imposta,

i condannati dalle varie Magistrature secondo la durata della pena. I coscritti secondo la statura, i censiti secondo l'età, ecc., formo, non delle serie, ma delle seriazioni. La scala può essere graduata mediante elementi dimensivi continui, come negli esempi ora citati, oppure mediante elementi discontinui, sebbene suscettivi di espressione numerica.

SERIE PERIODICHE (1).

MESI	Nati vivi in Italia nell'anno 1901 (periodo annuale)	Entrate e spese del bilancio dello Stato in Italia nell'esercizio 1898-1899 (in milioni di lire)			Press. barometrica diurna (solstizio d'estate)		Matrimoni per giorni della settimana a Cremona nel triennio 1887-1889	
		Date	Entrate	Spese	Ore del giorno	Altezza barom. oltre i 750 mm. (in mm.)	Giorni	Numero
Gennaio .	96.432	Luglio 1898	113,8	95,0	0	2,095	Lunedì . .	199
Febbraio .	92.590	Agosto »	155,9	103,2	2	1,686	Martedì . .	78
Marzo . .	100.459	Settembre »	108,2	88,2	4	1,229		
Aprile . .	90.996	Ottobre »	159,3	86,8	6	1,057	Mercoledì.	33
Maggio . .	87.962	Novembre »	107,1	76,2	8	1,337	Giovedì . .	123
Giugno . .	82.855	Dicembre »	230,3	416,2	10	1,658	Venerdì . .	7
Luglio . .	84.135	Gennaio 1899	127,9	94,4	12	1,822		
Agosto . .	85.030	Febbraio »	141,5	84,8	14	1,835	Sabato . .	101
Settembre	87.249	Marzo »	120,2	106,1	16	1,957	Domenica	185
Ottobre .	88.174	Aprile »	156,8	100,0	18	2,124		
Novembre	82.745	Maggio »	118,1	108,8	20	2,376		
Dicembre	79.136	Giugno »	217,8	411,1	22	2,342		

(1) È facile riconoscere che nella serie delle nascite il massimo apparente si ha nel marzo, mentre il massimo reale spetta al febbraio, perocchè, tenuto conto della diversa lunghezza dei due mesi, la media giornaliera in quello non arriva a 3241, in questo invece raggiunge la quota di 3307. Il minimo apparente dell'anno spetta poi al dicembre, con 2553 nascite giornaliere; ma è apparente soltanto, perchè, come vedremo, buon numero di nati a fin d'anno si denunciano ad arte come nati nel gennaio dell'anno successivo; sicchè il minimo reale compete forse al luglio (2714 nascite giornaliere).

La serie delle entrate e spese di bilancio presenta i massimi caratteristici del giugno e del dicembre, a causa soprattutto delle scadenze della rendita pubblica e della relativa ritenuta di ricchezza mobile; inoltre, quanto alle entrate, vi si distinguono dei massimi secondari, nei mesi pari, spiegabili colle scadenze bimestrali delle rate di imposte dirette.

L'onda barometrica diurna ha la sua maggiore altezza alle ore 20, la minore alle 6.

Caratteristico, infine, è il periodo settimanale della nuzialità. Sebbene i dati si riferiscano ad una piccola città, essi esprimono certo una situazione di cose piuttosto generale in Italia. Il noto pregiudizio del venerdì si rivela con un

Così se ripartisco gli alunni delle scuole per classi 1ª, 2ª, 3ª, ecc.; i biglietti emessi da una banca secondo i tagli (da 50 lire, da 100, da 500, ecc.); i telegrammi secondo il numero delle parole, il qual numero non può essere che intero, le seriazioni si diranno per elementi discontinui.

Esempi di seriazioni.

Inghilterra e Galles — Numero dei proprietari secondo l'estensione dei fondi

Acri	Proprietari nel 1873
meno di 1	703.289
1-10	121.983
10-50	72.640
50-100	25.839
100-500	32.317
500-1.000	4.799
1.000-2.000	2.719
2.000-5.000	1.815
5.000-10.000	581
10.000-20.000	223
20.000-50.000	66
Oltre i 50.000	4
	966.275

Biglietti consorziali definitivi e già consorziali di scorta in circolazione in Italia alla fine del 1882 e caduti in prescrizione il 13 aprile 1893

Tagli lire	In circolazione	Prescritti
0,50	22.223.942	2.876.572
1	38.986.857	2.115.531
2	31.988.696	477.477
5	40.000.000	243.204 [1]
10	24.000.000	85.343 [1]
20	2.500.000	9.239
100	600.000	3.617
250	336.320	1.092
1000	175.920	19
	160.811.735	5.812.094

Condannati nel 1901 in Italia dalle varie magistrature distinti secondo la durata della reclusione o detenzione

Durata della pena	Numero dei condannati
Fino a 6 mesi	142.236
Da 6 mesi a 3 anni	23.356
Da 3 a 5 anni	2.704
Da 5 a 10 anni	1.275
Da 10 a 20 anni	671
Oltre i 20 anni	304
Ergastolo	98
	170.644

Le seriazioni, come vedremo, descritte graficamente presentano forme alle volte assai regolari, che servono a classificarle. Abbiamo così seriazioni *simmetriche* e *asimmetriche*, *iperboliche*, *binomiali*, ecc., con *uno* o *più punti di massimo addensamento*, ecc.

Dei « prospetti » statistici in senso stretto non ci sembra il caso di recare esempi illustrativi. Facciamo eccezione per quelli che costituiscono un genere molto affine alle seriazioni propriamente dette; per quelli cioè in cui la frequenza del fenomeno è riferita ad elementi discontinui combinabili in un numero determinato di maniere. Supponiamo di dover analizzare la forma dell'esametro, nelle sue prime

---

minimum di nozze; il massimo sembra cadere in lunedì o domenica. Veggasi a questo proposito anche il risultato di un calcolo da noi istituito per tutta Italia, in base alle osservazioni del trentennio 1872-1901 (R. Benini, *Sul modo di ricavare la periodicità settimanale di un fenomeno, di cui son date le variazioni solo per mesi;* in *Giornale degli Economisti*, ottobre 1904).

(1) Caduti in prescrizione il 30 settembre 1894.

quattro sedi, presso un autore latino; è chiaro che bisognerà esaurire tutte le combinazioni possibili di *dattili* e di *spondei*, dalla forma estrema di quattro dattili a quella estrema di quattro spondei. Il Rasi, nel suo prezioso studio sull'*Arte metrica in Ennodio* (1), indica per 493 esametri di questo autore la seguente ripartizione:

FORME DELL'ESAMETRO NEI DISTICI ELEGIACI DI ENNODIO.

(*d* = dattilo; *s* = spondeo).

Con cominciamenti dattilici		Con cominciamenti spondaici	
dddd	11 esametri	*ssss*	41 esametri
ddds	31 »	*sssd*	10 »
ddss	83 »	*ssdd*	5 »
dsss	109 »	*sddd*	6 »
dsds	52 »	*sdsd*	12 »
ddsd	18 »	*ssds*	16 »
dssd	18 »	*sdds*	29 »
dsdd	12 »	*sdss*	40 »
Totale	334 esametri	*Totale*	159 esametri

Dove si nota come, nonostante la preponderanza delle voci spondaiche sulle dattiliche (1154 contro 818), che in generale caratterizza la *gravitas* della lingua latina, il poeta si attiene alla nota regola d'arte di dare un colorito, una fluidità maggiore al verso mediante il cominciamento dattilico. Una delle forme più frequenti appare la *ddss*, che nel verso epico, anche più che nell'elegiaco, si considerava squisita.

Altri esempi ci forniscono le combinazioni matrimoniali per stato civile, religione, nazionalità, professione, luogo di nascita, ecc., degli sposi.

---

(1) V. PIETRO RASI, *Dell'arte metrica di Magno Felice Ennodio, vescovo di Pavia* (*Bollettino della Società pavese di storia patria*, marzo-giugno 1902).

CAPO SECONDO

Critica e comparazione dei dati primitivi.

Sommario: § 1. Degli errori di rilevazione in generale. — § 2. Errori materiali. — § 3. L'*animus* dell'osservatore, come causa di errori. — § 4. Errori dovuti agli strumenti e modelli di rilevazione. — § 5. Errori dipendenti da ignoranza, diffidenza, ecc., degli interrogati. — § 6 e 7. Comparazione dei dati.

§ **1.** Primo e delicato ufficio della critica è quello di rintracciare gli errori, che possono essersi verificati nello stadio della rilevazione, misurarne la grandezza di effetto ed eliminarli, se possibile.

Il piano o programma della rilevazione, allo stesso modo che prestabilisce i limiti di tempo, di luogo, di specializzazione, ecc., deve prevedere gli errori temibili, dipendenti da qualunque causa: resistenze, diffidenze, ignoranza degli interrogati, inopportunità del momento, possibilità di interpretazioni discordi dei quesiti, negligenza degli incaricati delle operazioni di spoglio, ecc. Vi hanno dei limiti, difficilmente assegnabili, di tolleranza, oltre i quali è buon consiglio rinunziare all'indagine progettata. Degli errori, poi, altri si possono prevedere e prevenire, altri prevedere e non prevenire, ma soltanto correggere con metodi approssimativi, altri infine nè prevedere, nè prevenire, ma solo sospettare a rilevazione compiuta per via d'indizi, che la critica e la comparazione dei dati mettono in luce. La situazione naturalmente è diversa secondo che si tratta di rilievi *ex novo*, senz'altro precedente che l'osservazione comune o qualche assaggio statistico parziale, o invece si tratti di indagini continuative, che si giovano dell'esperienza formata mercè regolari rilievi nello stesso campo di esplorazione.

Gli errori prevedibili, si possano o non si possano prevenire, sono quelli in certa guisa virtualmente già conosciuti nelle loro cause; ed è appunto in ragion delle cause loro che si prestano ad una logica sistemazione della materia. Così si distingueranno gli errori dipendenti dallo stesso osservatore o dai suoi collaboratori o dagli interrogati, quelli dipendenti dagli strumenti e mezzi impiegati nell'indagine o dalla particolare natura del fenomeno, ecc. Ma con ciò la sistemazione non riesce completa; bisogna tener conto degli errori non prevedibili, ma solo riconoscibili o sospettabili a rilevazione compiuta. Il criterio discriminativo sarà qui fornito dal modo di riconoscimento. Noi vedremo per via di esempi pratici, più suggestivi delle formule astratte, che la presenza di un errore è da ritenersi probabile, quando le serie o seriazioni ottenute mostrano delle discontinuità, che

contraddicono alla continuità e regolarità naturale del fenomeno. altrimenti provata o ragionevolmente presunta. Del pari è fondato il sospetto di errore, quando il fatto rilevato in una certa fase discorda dal fatto stesso rilevato in una fase immediatamente successiva, ovvero discorda da fatti affini aventi con esso un rapporto ben stabilito di concomitanza o di conseguenza. Il dubbio può nascere pure dalla inverosimiglianza delle conclusioni, alle quali si andrebbe incontro, introducendo il dato rilevato come elemento di calcolo nella soluzione di speciali quesiti. E così dicasi di altri artifici di scoperta, in cui viene necessariamente in giuoco la comparazione dei dati.

Ora, se si pensa che la comparazione è una operazione essenzialmente logica e che l'errore è esso medesimo uno di quei fenomeni variabili cui si applicano i metodi statistici, apparirà chiaro che tutta questa parte critica, presupponendo la notizia delle regole dell'induzione statistica, dovrebbe avere la sua sede naturale tra gli ultimi capitoli dell'opera; tuttavia non esitiamo a collocarla qui, perchè in pratica l'esercizio della critica comincia dalla preparazione del programma e si spiega compiutamente a rilevazione ultimata, e perchè lo studioso, che ancora non conosca le regole dell'induzione statistica, si deve ritenere informato almeno di quelle dell'induzione ordinaria, le quali bastano per intendere ciò che verrà esposto nei seguenti paragrafi.

§ 2. Non dovremmo qui parlare di *errori materiali* di conteggio o di collocazione delle unità statistiche nelle caselle dei prospetti provvisori, presumendosi sufficiente ad evitarli una accurata revisione; ma poichè in qualche pubblicazione ufficiale ne sono rimasti alcuni senza *errata-corrige*, vale la pena di segnalarli a titolo d'esempio. La serie qui a fianco ci dà la percentuale degli atti di matrimonio sottoscritti da entrambi gli sposi in provincia di Perugia (compartimento dell'Umbria) negli anni dal 1866 al 1870 (1).

Umbria Anni	Atti sottoscritti da entrambi gli sposi
1866	12,22 per 100
1867	12,32 »
1868	22,99 »
1869	13,25 »
1870	15,02 »

Cotesta percentuale va soggetta d'anno in anno a così piccole oscillazioni, tanto nella provincia in questione quanto in ogni altra provincia italiana, che la cifra segnata pel 1868 (22,99 $^0/_0$!) deve essere per la sua eccezionalità ritenuta erronea. Non è improbabile che il

(1) *Movimento dello stato civile per l'anno* 1866 *e per i seguenti* (Direzione generale della Statistica).

compilatore abbia, senza completare lo spoglio delle schede, attribuito a tutta la provincia, per il 1868, le proporzioni trovate per la città capoluogo.

Lo stesso dicasi, ancora per l'Umbria, della classificazione delle spose per età nell'anno 1890 (1), la quale, confrontata con quella di qualsiasi altro anno, per lo stesso compartimento o per altri, presenta anomalie non spiegabili se non coll'ipotesi di un errore materiale di spoglio e d'aggruppamento.

Età delle spose	Numero nel 1890	Media del periodo 1888-1892 escluso il 1890
Meno di 17 anni	127	23
17-18 anni.	567	144
19-20 »	790	446
21-22 »	640	750
23-24 »	414	812
25-26 »	515	681
27-28 »	344	476
29-30 »	377	334
31-35 »	212	483
36-40 »	225	278
41-45 »	168	178
Oltre i 45 anni	220	203
	4.599	4.808

Queste anomalie sono: la eccezionale frequenza di spose men che ventenni; la superiorità numerica dei gruppi di 25-26 anni, di 29-30 e di 36-40 su quelli di 23-24, di 27-28 e di 31-35, rispettivamente, mentre l'opposto si verifica in ogni altro anno e in ogni altra parte d'Italia; la inferiorità numerica del gruppo 31-35, che comprende *cinque* classi annuali in confronto del gruppo 29-30, che ne comprende *due* sole; laddove qui pure si verifica costantemente il contrario. Infine il punto di massimo della seriazione, che suol cadere nell'intervallo dai 23 ai 24 anni, anticipa nel 1890 di un quadriennio. C'è n'è d'avanzo per sospettare la presenza di un errore. Ad ogni modo, se coteste anomalie fossero reali, è molto verosimile che ne avremmo di corrispondenti nella seriazione degli sposi per età secondo i dati del 1890; ma l'esame delle relative serie, riportate a pagina seguente, esclude affatto questa supposizione. Deve dunque trattarsi di uno di quegli errori materiali che un ben organizzato sistema di revisione tende a rendere sempre più rari nelle pubblicazioni ufficiali.

(1) *Movimento dello stato civile per l'anno* 1890 *e per i seguenti* (Direzione generale della Statistica).

Età degli sposi	Numero nel 1890	Media del periodo 1888-1892 escluso il 1890
Meno di 21 anni	44	48
21-22 anni	139	177
23-24 »	476	420
25-26 »	746	758
27-28 »	625	694
29-30 »	553	585
31-35 »	870	861
36-40 »	415	488
41-45 »	310	309
Oltre i 45 anni	421	468
	4.599	4.808

Un terzo caso di errore materiale abbiamo in una recente e preziosa monografia sugli *Effetti di commercio* in Italia, la quale indica, come si legge nel qui unito prospetto, la cifra di 6.556.374 effetti cambiari tassati in Italia nell'esercizio 1887-1888, cifra cui fa seguito con notevole sbalzo quella di 9.060.556 per il 1888-1889.

ESERCIZI	Numero degli effetti bollati	Valor medio in lire	Saggio della tassa per °/₀₀	Riscossione in lire
1887-1888	6.556.374	1876	0,60	7.379.019
1888-1889	9.060.556	1328	1,00	10.825.551
1889-1890	10.429.774	885	1,20	11.065.292
1890-1891	10.137.500	812	1,20	9.878.352
1891-1892	9.921.974	783	1,20	9.316.829
1892-1893	9.612.759	773	1,20	8.903.056

Il valor medio d'ogni effetto sarebbe poi precipitato da lire 1876 a 1328. Ora i dati del primo esercizio appaiono fortemente sospettabili d'errore per due motivi soprattutto: 1° perchè notoriamente il 1887-1888 fu epoca di massima espansione d'affari in Italia, tanto che allora gli sconti delle Banche di emissione e di altri Istituti di credito raggiunsero il loro più alto punto; 2° perchè nel 1887-1888 la tassa graduale di bollo sulle cambiali, al saggio di lire 0,60 per mille, rese ben 7.379.019 lire, mentre nell'esercizio successivo al saggio medio di lire 1,00 per mille ne rese 10.825.551; il che significa che al saggio di 0,60 avrebbe fruttato solo 6.495.331 lire. L'aumento di 40 centesimi di tassa su 1000 lire non potè avere un'influenza sensibile sull'uso delle cambiali, perchè aumenti ben più rilevanti nel saggio dello sconto si chiarirono in quel periodo di speculazione inefficaci a frenare la corsa allo sconto; e d'altronde, se se ne ebbe una, dovette essere nel senso di scemare il numero degli effetti creati

nel 1888-1889 in confronto del precedente esercizio. Sicchè molto probabilmente la massa delle cambiali create nel 1887-1888 sta a quella dell'esercizio successivo come lire 7.379.019 stanno a lire 6.495.331. E poichè nel 1888-1889 il numero degli effetti fu di 9.060.556, quello del 1887-1888 non potè essere inferiore a 10 milioni (cifra che arrotondiamo per altre considerazioni, che è inutile riferire), di un valore medio di lire 1230. Infine la riscossione del 1888-1889, essendo avvenuta al saggio medio di 1 $^0/_{00}$, corrisponde ad un imponibile di 10.825 milioni, che divisi per 9.060.556 effetti dànno come valor medio di ciascuno la cifra di 1195 lire. Gli sbalzi della serie risultano così molto attenuati e credibili.

L'errore materiale esisteva infatti e può dimostrarsi. Mentre il dato degli effetti bollati per il 1888-1889 comprende tanto le cambiali, in cui la tassa di quietanza è compenetrata nel prezzo della carta filigranata, quanto quelle scritte su altra carta, cui si applicavano poi per la quietanza apposite marche da bollo, il dato per il 1887-1888 tien conto delle prime solamente e non delle seconde, che nelle statistiche del tempo sono confuse con altri titoli di riscossione, da cui non si possono sceverare (1).

§ **3.** L'*animus*, l'elemento intenzionale dell'osservatore, come causa di errori nella rilevazione dei dati, non si presume, ma nemmeno si esclude, quando sia in giuoco una tesi, cui è moralmente interessato l'osservatore medesimo. C'è da meravigliarsi, per es., che un clinico citi piuttosto i casi di guarigione di un suo speciale metodo di cura e attribuisca i casi sfavorevoli a circostanze particolari, di cui tende ad esagerare l'importanza? No, certo. L'osservatore deve avere quella specie di coraggio che consiste nell'esercitare l'autocritica e nell'ammettere il controllo di interessi avversi. Gli organi collettivi, massime i pubblici, quantunque senza confronto men sospettabili, convien pure che si costituiscano in maniera da dare la rappresentanza ai varii interessi in contrasto. Così, un'inchiesta sulle condizioni igieniche delle abitazioni in un Comune potrebbe dare risultati differenti, se

---

(1) Ciò è chiaramente indicato a pag. 241, nota *d*, e a pag. 242, nota *c*, del volume: *Tasse sugli affari;* Dati statistici relativi alle riscossioni dal 1° luglio 1887 al 30 giugno 1888. Ministero delle finanze, Roma. Cfr. colla pubblicazione dell'esercizio successivo.

La monografia sugli *Effetti di commercio in Italia* è nel *Bollettino di statistica e legislazione comparata*, della Direzione generale del Demanio e Tasse sugli affari, anno 3° (1902-1903), fasc. 3°, pag. 470. Inutile dire che i nostri appunti, ispirati solo all'interesse della verità, la quale deve primeggiare su ogni altra considerazione, non scemano per nulla i meriti segnalati di questa pubblicazione e di quelle dianzi citate sul *Movimento dello stato civile*. Si tratta di errori accidentali e rarissimi, inevitabili anche nella miglior organizzazione delle statistiche ufficiali.

ispirata e condotta solo dai rappresentanti della classe proprietaria o solo da quelli della classe operaia; una rappresentanza mista è garanzia di sincerità e obbiettività dell'indagine.

Prescindendo dall'elemento intenzionale o quasi-intenzionale, si dànno errori derivanti dalla varia finezza dei sensi degli osservatori. Il principio o la fine dell'occultazione di un astro dietro un altro astro non sono percepiti nell'identico istante da diversi astronomi; chi vede o crede vedere un po' prima e chi un po' dopo. Per ogni osservatore si può determinare la cosidetta *equazione personale*. In certi casi converrebbe, come già ebbe a proporre lo Schiaparelli, istituire una specie di « rotazione » degli osservatori, affine di eliminare o compensare gli errori personali, per es., nelle letture delle altezze barometriche e simili.

È comunemente nota la simpatia per le cifre rotonde e l'avversione per i numeri dispari (eccetto il 5, che si considera come cifra rotonda), contratta forse in grazia della maggior comodità che le prime offrono nelle ordinarie operazioni aritmetiche. La si avverte così nelle misure antropometriche, come nelle letture dei decimi di grado sul termometro e come nelle assegnazioni di punti di merito negli esami scolastici.

TEMPERATURE REGISTRATE IN 28 OSSERVATORII (1).

Decimi oltre la cifra intera dei gradi	Massimi mensili	Minimi mensili	Totale
0	61 volte	74 volte	135 volte
1	13 »	21 »	34 »
2	36 »	36 »	72 »
3	24 »	27 »	51 »
4	31 »	22 »	53 »
5	34 »	31 »	65 »
6	40 »	26 »	66 »
7	19 »	25 »	44 »
8	39 »	39 »	78 »
9	26 »	22 »	48 »
Totali . . .	323 volte	323 volte	646 volte

Così nel presente prospetto troviamo che la lettura di un numero intero di gradi con 0 decimi si è ripetuta 61 volte su 323 massimi regi-

(1) *Annuario statistico italiano per l'anno* 1900, pag. 26. Manca per qualche mese e per qualche osservatorio il dato della temperatura; ciò spiega perchè si hanno 323 osservazioni invece di 336 (= 28 × 12).

strati e 74 volte su 323 minimi, ossia con una frequenza doppia di quella che le sarebbe spettata a parità di ogni altra circostanza; invece la lettura di un numero intero di gradi, più 1 decimo, si verificò solo 13 volte nella prima serie e 21 nella seconda; e quella di un numero intero più 7 decimi, solo 19 e 25 volte rispettivamente, ecc.

Ora, senza dubbio alcuno, non è che la natura aborra dalle temperature terminanti in 1, 3, 7 o 9 decimi e simpatizzi per le altre; la simpatia per le cifre rotonde e i numeri pari è negli osservatori medesimi.

In quest'altro prospetto, che concerne risultati di esami scolastici, debbono considerarsi rotondi, per ragioni che tutti sanno, il 30, il 27, il 24, il 21, il 18, eguali rispettivamente al triplo di 10, 9, 8, 7 e 6; ed è giusto in corrispondenza loro che si hanno i risultati più frequenti degli esami.

RISULTATI DI 3204 ESAMI SPECIALI (178 ESAMINATI) (1).

	N° degli esami
30 punti con lode	27
30 »	219
29 »	122
28 »	184
27 »	484
26 »	238
25 »	314
24 »	341
23 »	168
22 »	186
21 »	220
20 »	206
19 »	96
18 »	330
Meno di 18 punti	69
	3.204

A voler dare alla serie la sua probabile continuità e regolarità naturale, occorrerebbe trasformarla con qualcuno dei metodi di perequazione, dei quali terremo parola in sede opportuna.

Quando si tratti di numeri non primitivi, ma *derivati* (come prodotti o quozienti), le agglomerazioni originarie intorno alle cifre rotonde facilmente scompaiono nei dati derivati, ove anzi si possono dare, come conseguenza del calcolo aritmetico, agglomerazioni intorno alle cifre non rotonde. Supponiamo, ad es., che intorno alla cifra rotonda di 150 mm. per il diametro trasversale del capo e a quella pure rotonda di 190 mm. per il diametro longitudinale si siano regi-

(1) Da nostre ricerche sui risultati degli esami speciali nella Facoltà di giurisprudenza all'Università di Pavia.

strate più del bisogno le osservazioni; il rapporto $\frac{150 \times 100}{190} = 79$, che esprime l'indice cefalico, segnerà un'agglomerazione intorno ad una cifra non rotonda. Il Livi ha in tal maniera dimostrato che gli indici di 79, 84, 86 e 89 sono più frequenti del verosimile e quelli di 80, 85 e 90 meno frequenti, in apparenza contro la regola della simpatia per le cifre rotonde, in realtà per un semplice giuoco di aritmetica (1).

§ 4. Una parte non minore nella comparsa degli errori spetta di solito agli strumenti e ai modelli della rilevazione. I questionari, che non contengano istruzioni brevi, chiare, precise per l'intelligenza delle domande proposte, sono da riprovarsi. Se, poniamo, in un censimento delle case si trascura di precisare il significato del termine « casa » avverrà facilmente che da taluno s'intenda la casa come corpo di fabbricato unico, per sè stante dal punto di vista architettonico, indipendentemente dal fatto che esso sia posseduto per distinte porzioni da diversi proprietari; da altri invece si guardi piuttosto al fatto economico-giuridico della proprietà. I modelli di rilevazione troppo complicati espongono al pericolo di risposte date a caso e benanco alla renitenza a rispondere.

Nelle scienze fisiche l'interesse a correggere gli errori derivanti dall'imperfezione degli strumenti è tanto più grande quanto minore è il numero delle osservazioni che si possono o si vogliono ripetere sullo stesso oggetto, ed è tanto più grande quanto più gravi si presumono le conseguenze teoriche o pratiche di rilievi insufficientemente approssimati. I procedimenti di correzione o eliminazione variano poi a seconda dei casi e degli strumenti. Così l'errore proveniente dalla imperfezione del tamburo e della vite micrometrica dei microscopi si elimina col far collimare il filo del microscopio colla divisione immediatamente a destra e con quella immediatamente a sinistra e prendendo la media delle letture così ottenute. L'errore di eccentricità del cannocchiale si elimina col puntare ogni oggetto con il cannocchiale a destra e a sinistra, ecc. Peraltro, nelle osservazioni che non pretendono ad una estrema precisione o che per il loro grande numero debbano essere sbrigate con rapidità, ci contentiamo spesso di sapere che l'errore dello strumento è in eccesso anzichè in difetto, o viceversa, e che non oltrepassa certi limiti. Nelle misure dei diametri del capo sui nostri arruolati, i medici militari fecero uso del quadro a massima, che dà di solito valori più grandi del compasso, come quello che non viene a perfetto contatto colla cute, trovando ostacolo nei capelli.

---

(1) V. *Antropometria militare*, pag. 80 e 81.

La natura stessa dell'oggetto da rilevare o le circostanze momentaneamente sfavorevoli all'osservazione esigono pure attento esame. I movimenti d'inspirazione e di espirazione rendono più o meno inesatta la misura del perimetro toracico in una prima ed unica prova. Un censimento di sordomuti riesce dubbio per le prime età, molti bambini, affetti davvero da sordomutismo, essendo ritenuti dai loro parenti tardivi all'acquisizione della parola. Subito dopo una battaglia, il conto dei morti, feriti, prigionieri e dispersi non può riuscire che approssimativo.

§ **5.** La categoria più larga di errori è però quella che trae origine dall'ignoranza, dalla diffidenza o dal malinteso interesse degli interrogati. Le statistiche della produzione agraria e industriale lasciano notoriamente molto a desiderare. Le denunzie dei redditi e quelle dei valori ricevuti per eredità o donazione sono d'un bel tratto al di sotto del vero. Si calcola, per esempio, che circa la metà dei titoli pubblici, azioni e obbligazioni di Società e simili, trasferiti *mortis causa*, sfugga al fisco italiano (1). Ma l'interesse è talvolta anche soltanto morale. Così, nei censimenti in cui si domanda se della famiglia fanno parte individui idioti, cretini o pazzi, il quesito ottiene risposte incomplete; molte famiglie non amano mettere in pubblico disgrazie che le affliggono e le umiliano. Per analogo motivo, individui morti di malattie tubercolari son dichiarati morti di bronchite o di polmonite cronica. Nei paesi cattolici, per riguardi religiosi, un certo numero di suicidii si fan passare come morti naturali. E si potrebbero moltiplicare gli esempi, chè quasi ogni statistica ne offre più d'uno; tuttavia ci limiteremo a trattare un po' diffusamente di un caso caratteristico.

---

(1) I titoli emessi dallo Stato e da società ed enti diversi dallo Stato possono calcolarsi in 17 miliardi e mezzo di consistenza attuale. Deducendo i titoli collocati all'estero (circa 2 miliardi, di cui non vi può essere traccia nelle successioni, che si aprono in paese), la rendita nominativa intestata allo Stato, a Corpi morali, ecc., insomma a persone giuridiche, che non muoiono come le persone fisiche (altri 3 miliardi e un quarto), i titoli al portatore di proprietà di società e istituti (2 miliardi?), rimangono in cifra tonda 10 miliardi come valor nominale di titoli posseduti da privati. Di tanta somma, una quarantesima parte circa, a nostro avviso, dovrebbe figurare nelle successioni di ogni anno (donazioni comprese, in quanto si considerano come semplici anticipi di eredità), perchè ogni anno muore un quarantesimo della popolazione maggiorenne, tra cui sono generalmente diffusi tali titoli, e sottostare all'imposta. In breve, il fisco dovrebbe ritrovare in detti trasferimenti una somma imponibile non lontana dai 250 milioni annui; invece non riesce a sorprendervi che 135 milioni in media! È quanto dire che il 47 % dei titoli in parola sfugge all'imposta di successione. Questa proporzione crescerebbe ancora se considerassimo a parte i soli titoli al portatore (V. *Bollettino di statistica e legislaz. comparata*, della Direzione generale del Demanio e Tasse sugli affari, anno 1°, vol. I, pag. 777).

In Sicilia e in gran parte dell'Italia meridionale, il rapporto delle nascite maschili alle femminili, *così poco variabile negli altri mesi dell'anno*, cambia bruscamente nel dicembre e nel gennaio; in gennaio troppe nascite maschili in paragone delle femminili; in dicembre troppo poche. Ciò deve imputarsi con tutta verosimiglianza al costume di denunziare i maschi, se nati negli ultimi giorni dell'anno, come nati nei primi dell'anno successivo, allo scopo di far loro ritardare di un anno la leva (1). Il fenomeno sembra risalire infatti all'epoca in cui entrò in vigore la legge 7 giugno 1875 sul reclutamento. Anche qui si tratterebbe dunque di un interesse non propriamente economico.

SICILIA

MESI	MASCHI PER 100 FEMMINE		
	1895	1896	1897
Gennaio	**118,82**	**118,09**	**125,62**
Febbraio	107,72	105,51	103,00
Marzo	103,06	104,03	104,53
Aprile	106,95	106,44	107,58
Maggio	101,25	105,08	110,62
Giugno	106,12	106,71	103,75
Luglio	107,19	106,01	107,39
Agosto	104,81	108,07	101,56
Settembre	111,76	100,37	107,26
Ottobre	104,90	105,04	109,02
Novembre	103,90	101,11	102,15
Dicembre	**90,42**	**86,58**	**88,41**
Media . . .	105,63	104,73	105,90

La serie dianzi data per il triennio 1895-1897 insegna che mentre negli altri mesi dell'anno il rapporto non si è distaccato dalla media che di 6 punti al massimo, nel dicembre fu invece inferiore di almeno 15 punti e nel gennaio superiore di almeno 13. Ad ogni modo per dissipare il dubbio che si tratti di accidentalità, ci piace rimontare con la serie al 1880; apparirà chiaro che il fenomeno è costante e si va anzi accentuando col tempo.

(1) In via approssimativa calcolo per tutta Italia un 2500 maschi, indebitamente iscritti ogni anno tra i nati del gennaio, mentre dovrebbero figurare tra quelli del dicembre antecedente. Non è poi escluso che in qualche misura il ritardo delle denunzie avvenga anche per le femmine, naturalmente per motivo diverso da quello segnalato pei maschi.

NATI MASCHI PER 100 FEMMINE IN SICILIA

ANNI	Gennaio	Dicembre	Media dell'anno	Eccedenza in gennaio	Deficienza in dicembre
1880	113,60	100,33	106,29	+ 7,31	— 5,96
1881	109,72	98,58	105,87	+ 3,85	— 7,29
1882	110,89	100,51	106,03	+ 4,86	— 5,52
1883	109,71	100,34	106,00	+ 3,71	— 5,66
1884	116,74	97,29	106,02	+ 10,72	— 8,73
1885	111,28	98,04	106,07	+ 5,21	— 8,03
1886	112,56	100,47	106,01	+ 6,55	— 5,54
1887	119,04	98,47	106,19	+ 12,85	— 7,72
1888	115,82	94,53	105,66	+ 10,16	— 11,13
1889	115,60	94,71	105,71	+ 9,89	— 11,00
1890	115,47	93,47	105,10	+ 10,37	— 11,63
1891-1894 (1). . .	?	?	—	?	?
1895	118,82	90,42	105,63	+ 13,19	— 15,21
1896	118,09	86,58	104,73	+ 13,36	— 18,15
1897	125,62	88,41	105,90	+ 19,72	— 17,49
1898	117,21	86,72	105,98	+ 11,23	— 19,26
1899	119,16	86,22	105,96	+ 13,20	— 19,74
1900	122,15	88,26	105,80	+ 16,35	— 17,54
1901	120,25	87,31	105,07	+ 15,18	— 17,76
1902	121,25	86,48	107,07	+ 14,18	— 20,59

Se la causa da noi supposta è la vera, il ritardo delle denunzie dei maschi deve riguardare soprattutto i nati legittimi, poichè l'interesse per l'avvenire del figlio legittimo la vince certamente e di gran lunga su quello che si può avere per l'illegittimo. Ora anche questa illazione trova una piena conferma nelle cifre. Nei mesi di gennaio e dicembre del settennio 1884-1890 si ebbero complessivamente in Sicilia 175.364 nascite (esclusi sempre i nati-morti) così ripartite:

	GENNAIO		DICEMBRE		MASCHI per 100 femmine	
	Maschi	Femmine	Maschi	Femmine	Gennaio	Dicembre
Legittimi	48.578	42.007	35.643	36.999	115,64	96,34
Illegittimi ed esposti. .	3.388	3.112	2.851	2.786	108,87	102,33

(1) Per questi anni manca nelle pubblicazioni del *Movimento dello stato civile* la distinzione per sesso nei singoli mesi; come dal 1895 in poi manca la distinzione secondo l'origine legittima o illegittima dei nati nei singoli mesi.

Il contrasto delle percentuali del gennaio e del dicembre è, come prevedevamo, assai più accentuato tra i nati legittimi; esso sorpassa i 19 punti; tra gli illegittimi non arriva a 7 punti.

Vediamo qualche altro caso caratteristico.

Effetto di ignoranza o incuria degli interrogati, in occasione di censimenti, è il loro agglomerarsi intorno alle età rotonde, cioè terminanti per *zero* o per *cinque*, a scapito, ben s'intende, delle età circostanti. È un fatto consimile a quello segnalato nelle misure antropometriche, termometriche, ecc., ma determinato da cause speciali. Così si hanno, apparentemente, pochi individui di 29 anni o di 31, ma molti di 30; pochi di 39 o 41, ma molti di 40; e via dicendo (1). Che queste inesatte denunzie di età dipendano dall'ignoranza, in cui sono molti del popolo per ciò che riguarda la propria data di nascita, si argomenta dal fatto che tali agglomerazioni intorno alle età rotonde sono di gran lunga più rilevanti tra la popolazione analfabeta, che non tra quella che sa leggere e scrivere, più tra gli abitanti delle campagne che tra quelli delle città, più tra i meridionali che tra i settentrionali. L'abitudine per parte di alcuni di contare l'anno incominciato in luogo di quello compiuto è bensì una causa di errore, ma non producente agglomerazioni intorno alle età rotonde. Non

(1) Che le agglomerazioni in corrispondenza delle età rotonde si formino a scapito delle età circostanti, si vede subito dai seguenti dati del censimento italiano del 1881, riguardanti la popolazione maschile dei capoluoghi di provincia:

Età	Maschi censiti	Età	Maschi censiti	Età	Maschi censiti	Età	Maschi censiti
27-28	34,664	37-38	30,301	47-48	22,627	57-58	16,845
28-29	36,025	38-39	31.121	48-49	23,708	58-59	16,189
29-30	32,626	39-40	27,075	49-50	19,124	59-60	14,359
30-31	**43,241**	**40-41**	**41,585**	**50-51**	**36,524**	**60-61**	**27,962**
31-32	33,516	41-42	30,272	51-52	22,264	61-62	16,677
32-33	32,304	42-43	30,424	52-53	22,122	62-63	15,166
33-34	32,966	43-44	24,814	53-54	19,607	63-64	12,593
34-35	30,330	44-45	26,005	54-55	20,089	64-65	11,922
35-36	**34,216**	**45-46**	**29,478**	**55-56**	**20,873**	**65-66**	**12,049**
36-37	32,871	46-47	24,368	56-57	20,270	66-67	10,721

La semplice ispezione di queste cifre mostra una agglomerazione *principale* nelle età terminanti per zero (come tali debbono considerarsi quelle da 30 a 31, da 40 a 41, ecc.), e una *secondaria* nelle età terminanti per cinque (come tali si considerino quelle da 35 a 36, da 45 a 46, ecc.). È verosimile che l'attrazione delle prime si estenda fino a due anni e mezzo avanti e due anni e mezzo dopo l'età rotonda; e quella delle seconde fino ad un anno e mezzo avanti e ad un anno e mezzo dopo. Perciò convien perequare le prime sommando la classe dell'età rotonda colle due precedenti e le due susseguenti, più metà della terza antecedente e della terza susseguente e dividendo il risultato per 6; e convien perequare le seconde dividendo per 4 la somma della classe della età rotonda, più le due immediatamente contigue, più metà di ciascuna delle due contigue a queste.

devesi poi tacere, almeno per ridurre ne' suoi giusti confini l'opinione corrente, che le stesse donne, le quali rifuggirebbero più degli uomini dall'indicare la « decina » compiuta, si agglomerano invece intorno alle età terminanti per zero con singolare preferenza; la qual cosa si spiega facilmente se si riflette essere di solito l'uomo che nella sua qualità di capo-famiglia, anche senza consultare la donna, riempie per lei la scheda o detta le risposte al commesso del censimento. D'altronde fra le classi lavoratrici di città e di campagna costituenti la gran maggioranza della popolazione, la donna poco si preoccupa di nascondere la propria età e di evitare la decina compiuta. Infine è degno di nota il fatto che l'agglomerazione dei censiti cresce per ambo i sessi dai 30 ai 60 anni, cioè quanto più ci allontaniamo da avvenimenti caratteristici della vita, come la coscrizione, le nozze, ecc.; ma indi diminuisce, forse perchè per molti vecchi pensionati o ricoverati in ospizi di mendicità, di cronici e simili, l'assegnazione della pensione o l'accettazione nei ricoveri, subordinate al raggiungimento d'una data età, costituiscono di nuovo punti di riferimento vicini e facili pel calcolo agli stessi censiti o a chi s'incarica per loro di riempire la scheda.

Questo genere di errori si può eliminare, in via d'approssimazione, con processi elementari di perequazione. Senza addentrarci molto in un argomento che sarà svolto in apposita sede, ci limitiamo a dire, per l'intelligenza del prospetto seguente, che il numero probabile degli individui aventi proprio l'età terminante per zero, si è fatto eguale alla sesta parte del gruppo, che comprende l'età rotonda e i due anni e mezzo precedenti e susseguenti. Così la cifra perequata dei quarantenni *analfabeti*, che il prospetto indica in 12.286, si è ottenuta sommando le classi di 38, 39, 40, 41 e 42 anni più la metà della classe di 37 e metà di quella di 43 e dividendo il risultato per 6. Lo spediente involge l'ipotesi, accettabile come prima approssimazione, che entro certi limiti la popolazione decresca regolarmente di quantità eguali per ogni anno d'età. Ora, in confronto delle cifre perequate con il metodo anzidetto, noi daremo quelle greggie del censimento del 1881, restringendoci ai capoluoghi di provincia; ma distinguendo tra maschi e femmine, analfabeti e non analfabeti:

Anni d'età	MASCHI				FEMMINE			
	Analfabeti		Non analfabeti		Analfabete		Non analfabete	
	perequati	censiti	*perequati*	censiti	*perequate*	censite	*perequate*	censite
30	**12.190**	17.647	**23.065**	25.594	**17.216**	25.766	**18.568**	20.568
40	**12.286**	20.269	**19.053**	21.316	**16.734**	29.228	**14.179**	16.269
50	**10.363**	19.640	**13.780**	16.884	**14.537**	29.041	**9.911**	12.689
60	**7.913**	16.073	**9.599**	11.889	**11.226**	23.485	**6.943**	9.633
70	**3.721**	6.981	**4.699**	5.330	**5.584**	11.199	**3.407**	4.289

Ed ecco le percentuali d'errore, che si ricavano da questo prospetto:

Anni d'età	MASCHI		FEMMINE	
	Analfabeti	Non analfabeti	Analfabete	Non analfabete
30	30,9 %	9,9 %	33,2 %	9,7 %
40	39,4 »	10,6 »	42,7 »	12,8 »
50	47,2	18,4 »	49,9 »	21,9 »
60	50,8 »	19,3 »	52,2 »	27,9 »
70	46,7 »	11,8 »	50,1 »	20,6 »

Per *percentuale d'errore* o *di agglomerazione* intendesi il rapporto fra l'eccedenza dei censiti sui perequati e la cifra dei censiti fatta eguale a 100.

Il prospetto delle percentuali conferma quanto dicevamo sopra, cioè che questo genere di errori è in rapporto col sesso, coll'età e col grado di coltura. Il censimento del 1871 ha dato risultati analoghi, con percentuali però più alte.

Lo stato civile delle persone non è neppur esso senza influenza sul grado di agglomerazione dei censiti intorno alle età rotonde. In entrambi i censimenti del 1871 e 1881 le vedove diedero agglomerazioni assai più rilevanti che non le nubili o le coniugate coetanee, forse perchè, vivendo isolate, non hanno chi le aiuti nel richiamare le date caratteristiche della vita (1).

(1) Nel censimento ultimo nostro, del 9 al 10 febbraio 1901, essendosi richiesto, non più il dato diretto dell'età, ma quello dell'anno e mese di nascita, le agglomerazioni dovettero probabilmente verificarsi non già in corrispondenza delle età rotonde, ma delle date rotonde di nascita, come il 1830, il 1840, il 1850. Senonchè i risultati resi pubblici per tutto il regno figurano raccolti in gruppi quinquennali, quadriennali e triennali per le date di nascita anteriori al 1886, il che impedisce che si abbia la riprova della nostra supposizione. Un indizio ce lo fornisce tuttavia la pubblicazione del Municipio di Milano (*La popolazione di Milano secondo il censimento eseguito il 9 febbraio* 1901, Milano, E. Reggiani, 1903, pag. 13 e 14), nella quale appunto si rileva il numero eccezionale di censiti che indicarono una data rotonda di nascita, ovvero (ciò che è pure interessante) una data storica caratteristica, esempio, anni di rivoluzione o di guerra. Veggansi, infatti, i seguenti dati:

Censiti in Milano nel 1901, che si dichiararono nati

nel 1829	1367	nel 1839	3202	nel 1849	4547
» 1830	1893	» 1840	4025	» 1850	5382
» 1831	1763	» 1841	3629	» 1851	5206

e questi altri, che mettono in risalto la simpatia per le date memorande. 1848 (rivoluzione di Milano e guerra coll'Austria), 1859 e 1866 (anni di guerra coll'Austria).

Censiti in Milano nel 1901, che si dichiararono nati

nel 1847	4185	nel 1858	6042	nel 1865	7484
» 1848	5776	» 1859	6986	» 1866	8064
» 1849	4547	» 1860	6402	» 1867	7671

§ **6.** Negli esempi svolti sin qui noi abbiamo esplicitamente o implicitamente fatto ricorso all'operazione logica della comparazione. Altra via non si conosce per iscoprire gli errori di rilevazione. *Comparazione* è in sostanza la critica stessa in quanto mira a stabilire i limiti di omogeneità di due dati statistici, di cui uno si voglia far servire di controllo all'altro e a mettere in evidenza gli elementi intrinseci differenziali, che spiegano in parte la loro diversità quantitativa, e gli elementi estrinseci (vizi di rilevazione e simili), da cui deriva l'ulteriore loro discordanza.

L'*errore* è un caso particolare della discordanza; si potrebbe definire *una discordanza non necessaria.*

La scelta degli elementi, che servono come pietra di paragone della attendibilità di un rilievo statistico, obbedisce a norme, le quali, senza aver nulla di assoluto, semplificano però il còmpito della critica. Per orientarsi in questa scelta bisogna esaurire tutto un ordine metodico di ricerche.

Anzitutto è da vedere se il fatto in esame fu rilevato altre volte in condizioni simili o dissimili. All'uopo si passeranno in rassegna i punti di somiglianza e di differenza delle varie rilevazioni, a cominciare dal modo con cui si è definito l'oggetto nell'una e nelle altre, venendo via via alle caratteristiche speciali di tempo, di luogo, ecc. Non sarà mai abbastanza raccomandato allo statistico di esaminare se i fatti paragonati si sono svolti sotto l'influenza di disposizioni legislative uniformi (ben inteso, quando si tratti di fenomeni sociali su cui possono avere presa le leggi positive) e vennero rilevati con metodi uguali. Il confronto della nuzialità italiana subito prima e subito dopo il 1° gennaio 1866 svierebbe facilmente le conclusioni di coloro i quali dimenticassero che quella fu la data dell'introduzione del matrimonio civile, riforma accolta non senza resistenza, massime dalle popolazioni già soggette al Governo pontificio. Così riuscirebbe men che corretto il giudizio comparativo della nati-mortalità italiana e della francese, se non tenessimo presente che da noi si considerano nati-morti i bambini morti prima o durante il parto, laddove in Francia si registrano nella rubrica *mort-nés* tutti i bambini *présentés sans vie* all'ufficiale di stato civile e quindi bene spesso dei bambini che nacquero vivi e vitali, ma soccombettero per accidentalità nelle prime ore o nei primissimi giorni dalla nascita, il termine per la presentazione allo stato civile essendo in Francia di tre giorni. Insomma bisogna saper sceverare la discordanza necessaria — imputabile nel caso concreto alle differenze legislative, oltre che alle differenze economiche e sociali dei due popoli — dalla discordanza non necessaria, cioè dall'errore.

In secondo luogo può darsi che lo stesso fatto si trovi rilevato in due fasi immediatamente successive, quasi simultanee, sicchè teoricamente, ossia nell'ipotesi di rilevazioni perfette, le due statistiche dovrebbero collimare punto per punto. Se non collimano, sorge la legittima presunzione dell'errore, del quale resterà a trovarsi la sede e la misura

almeno approssimativa. Valga anche qui un paio di esempi tratti dalle statistiche del commercio internazionale e da quelle dell'emigrazione.

Teoricamente, alle quantità e ai valori delle merci registrate alla esportazione da un paese per un altro, dovrebbero corrispondere le quantità e i valori registrati all'importazione dal paese ricevente. Il fenomeno è il medesimo in due fasi quasi simultanee; l'omogeneità intrinseca non potrebbe essere più grande. Tuttavia la concordanza delle statistiche doganali non si avvera mai (1), per ragioni che brevemente riassumiamo:

1° il *cambiamento di destinazione dei carichi viaggianti*. Sia un carico di grano a Odessa dichiarato in esportazione per la Francia; nel viaggio il capitano della nave riceve ordine di sbarcare il grano a Genova. Le statistiche russa, francese e italiana non possono più essere d'accordo;

2° la presenza di *depositi franchi* e *magazzini generali*, ove entrano le merci in attesa di definitiva destinazione, che può essere diversa da quella indicata alla partenza dal paese d'origine;

3° le *denunzie generiche* delle merci d'esportazione, che la dogana del paese esportatore ha poco interesse a verificare con esattezza, trattandosi per lo più di prodotti esenti da dazio d'uscita; mentre la dogana del paese importatore, che li colpisce di dazio, è interessata a specificarne la qualità e il valore. Dicasi lo stesso dell'abitudine, che hanno molti commercianti, di indicare come *paese di destinazione definitiva* quello che è semplicemente *paese di transito;*

4° i diversi criteri per la *determinazione* dei *valori doganali*. In qualche paese si sta alle dichiarazioni dei commercianti; altrove, come da noi, il còmpito è affidato ad apposite Commissioni che, interrogate

---

(1) Veggasi il confronto delle statistiche italiane e francesi concernenti gli scambi dei due paesi:

Medie annuali, in milioni di lire.

PERIODI	Merci importate in Italia dalla Francia		Merci esportate dall'Italia in Francia	
	secondo la statistica italiana	secondo la statistica francese	secondo la statistica italiana	secondo la statistica francese
1879-1881	322,9	190,6	509,0	396,7
1882-1884	358,0	183,0	461,3	385,8
1885-1887	334,9	187,3	455,0	293,2
1888-1890	162,0	137,7	165,3	145,5
1891-1893	157,1	128,9	148,3	135,8
1894-1896	142,1	115,9	144,5	121,1
1897-1899	143,2	162,3	154,5	143,6

La grande diminuzione degli scambi dal triennio 1885-1887 al 1888-1890 fu conseguenza ben nota dell'applicazione di tariffe differenziali o di rappresaglia, sostituite poi dalle tariffe generali fino al 1899

le Camere di commercio, le principali Ditte, ecc., stabiliscono i valori medii dell'annata per merci singole o per gruppi di merci molto affini. I raggruppamenti e le specificazioni di voci variano poi tanto nelle tariffe doganali dei diversi Stati, che la concordanza caso per caso è impossibile;

5° l'elemento perturbatore delle *spese di trasporto*, quando il paese esportatore e quello importatore non sono limitrofi o lo sono soltanto in parte. Infatti il paese esportatore comprende nel valore delle merci esportate solo le spese di trasporto dal luogo di produzione *al proprio confine*, e non già quelle *oltre il confine;* l'importatore invece calcola la merce entrata pel valore che essa ha dopo aver percorsi paesi di transito, se ci sono, o superati tratti di mare, pagando quindi tariffe ferroviarie, diritti di transito, noli, commissioni, ecc.;

6° il *contrabbando*. Una merce dichiarata, per es., in Germania all'esportazione per l'Italia, può arrestarsi in Isvizzera e di qui venire introdotta per contrabbando in Italia. La nostra dogana non registrerà una quantità e un valore d'importazione corrispondenti a quelli segnati all'esportazione dalla dogana tedesca.

Aggiungasi alle predette cause la diversità dei *metodi di calcolare le tare* (tal merce si registra in un paese, poniamo, a peso netto reale; in un altro a peso lordo; in un terzo a peso netto legale, cioè con deduzione di un tanto prestabilito per cento a titolo d'imballaggio o recipienti); la d*ifferenza di tempo* fra l'uscita della merce da un paese e la entrata nell'altro (esempio: un carico partito nella seconda quindicina di dicembre dal Plata o dall'Australia arriva in Europa a metà gennaio o anche più tardi; le registrazioni potrebbero concordare, ma son comprese in due annate diverse nelle statistiche dei paesi speditori e dei riceventi); le perdite per naufragi, avarìe, deperimenti durante il trasporto, ecc. (1).

Pur senza entrare in minute particolarità, si comprende come la rilevazione del paese importatore costituisca di regola un dato più attendibile di quella del paese esportatore, vuoi per la maggior precisione e il controllo più severo delle denunzie, vuoi per il carattere definitivo della fase in cui è entrato il fenomeno, il cambiamento di destinazione non essendo più facile alla merce sdoganata, come quando essa viaggiava sul mare o traversava un paese di transito.

Qualche cosa di simile può ripetersi dei movimenti migratorii. Qui pure, teoricamente parlando, al numero e specie di emigranti da un paese dovrebbero fare esatto riscontro il numero e la specie di immigranti nei paesi d'arrivo. Nel fatto la corrispondenza è un *desideratum*, che non si avvera mai. Oltre alla diversità dei metodi di accertamento

(1) Veggasi L. Bodio, *Relazione presentata al Consiglio del commercio* (*Atti del Consiglio dell' industria e del commercio*, sessione giugno 1893). Pubblicata dal Ministero d'agricoltura, ecc., Roma 1893.

da paese a paese, si dànno trasformazioni o deformazioni del fenomeno da una fase alla successiva. Un emigrante italiano, ad esempio, che ha dichiarato di voler imbarcarsi all'Havre per gli Stati Uniti, trova lavoro in Francia e qui si ferma; un altro, che ha dichiarato d'emigrare temporaneamente, fissa invece una residenza *sine die* nel paese già designato o in un terzo, ecc.

La comparazione può istituirsi ancora tra un fatto e un altro fatto che abbia con esso un rapporto conosciuto di corrispondenza, di concomitanza, di conseguenza. Così la statistica del cambio sull'estero tratta dai listini delle Borse italiane trova il suo naturale termine di confronto in quella del cambio sull'Italia alle Borse di Parigi, di Londra, ecc. La statistica delle lettere affidate alla posta lo trova in quella dei francobolli più comunemente usati nell'affrancazione delle corrispondenze ordinarie (1). La statistica degli sposi, che non hanno saputo apporre la propria firma all'atto di matrimonio, si confronterà con quella dei coscritti analfabeti, e via dicendo.

Si può ancora uscire dall'àmbito di fatti specifici e concreti e assumere come termine di paragone un principio induttivo, una generalizzazione di fatti. Se il consumo di una merce apparisse diminuito, mentre i prezzi sono discesi, avremmo un discordanza col principio generalmente accettato, che cioè il consumo cresce e non diminuisce col ribassare del prezzo; all'analisi spetterebbe stabilire se la discordanza è reale e prodotta da altra causa operante in senso contrario dei prezzi o se invece nasconde un semplice errore di rilevazione. Del pari se trovassimo registrata una serie di giorni sereni concomitante con una grande depressione barometrica, l'inverosimiglianza balzerebbe all'occhio; o il primo dato o il secondo dev'essere errato.

La presunzione della continuità e regolarità di certe serie rientra nella sfera di coteste generalità induttive, che adempiono bene all'ufficio di termini di confronto. Il lettore non avrà dimenticato che noi ce ne siamo già valsi per mettere in evidenza gli errori materiali della classificazione per età degli sposi in provincia di Perugia per l'anno 1890 e della proporzione degli atti di matrimonio sottoscritti da entrambi i coniugi, pure in detta provincia, nel 1868, nonchè le incomplete denunzie delle nascite maschili a fin d'anno in Sicilia e le agglomerazioni indebite di censiti in corrispondenza delle età rotonde in tutto il regno.

La delicata materia della comparazione ammette anche quest'altra

(1) Nell'esercizio 1886-1887 si avvertì che il numero delle lettere semplicemente affrancate o raccomandate, calcolate in base ad un accertamento ristretto a pochi giorni dell'anno, era superiore a quello dei francobolli venduti, da 5 centesimi in su, mentre è risaputo che molte lettere portano più di un francobollo e che, d'altra parte, molti francobolli di quell'ammontare si usano pure nella spedizione di stampati. Quest'evidente contraddizione di risultati decise ad inaugurare coll'esercizio 1887-1888 un metodo meno imperfetto di statistica postale.

forma di sistemazione, che è suggerita dalla quantità dei termini paragonati. Si possono confrontare infatti termini risultanti entrambi unicamente da rilevazione diretta, o termini, di cui l'uno provenga da rilevazione diretta e l'altro da un calcolo fondato alla sua volta su elementi di osservazione statistica; e perfino termini derivanti entrambi da calcolo o induzione. Dei due, quello che si ha motivo di riguardare come più verosimile, serve di controllo all'altro.

a) Comparazione tra fatti risultanti entrambi da una rilevazione propria o diretta. — Non ci pare il caso di addurre esempi, dopo quanto si è detto a proposito delle statistiche doganali, di quelle dell'emigrazione, del cambio sull'estero, ecc.

b) Comparazione tra un fatto statisticamente rilevato e uno calcolato. — Agli esempi ricordati della classificazione effettiva dei censiti per età confrontata colle risultanze di un calcolo di perequazione, ecc., non sarà inutile aggiungere il seguente:

Il censimento nostro del 1881 ha dato un numero di bambini certo inferiore al vero. L'errore si è facilmente riconosciuto confrontando i censiti delle prime classi d'età coi bambini calcolati superstiti dalle nascite dell'anno stesso che si chiuse col censimento, dalle nascite dell'anno avanti, di due anni avanti, ecc., depurate delle morti avvenute nell'intervallo. Per le particolarità del metodo rimandiamo tuttavia al lavoro compiuto dall'Ufficio matematico di Statistica presso la nostra Direzione generale (1); qui ci limitiamo a considerare le risultanze del confronto. Così invece di 791.699 censiti in età da 0 a 1 anno, se ne dovevano avere 936,623 e invece di 611.294 censiti in età da 1 a 2 anni, se ne dovevano avere 709.322. Le classi successive presentano discordanze molto meno rilevanti.

Età	Ambo i sessi		Differenze
	Censiti	Calcolati	
0-1 anno	791.699	936.623	+ 144.924
1-2 anni	611.294	709.322	+ 98.028
2-3 »	706.630	722.708	+ 16.078
3-4 »	677.559	663.227	— 14.332
4-5 »	652.421	653.877	+ 1.456
5-6 »	660.770	670.957	+ 10.187
6-7 »	646.659	632.872	— 13.787
7-8 »	589.283	574.348	— 14.935
8-9 »	572.187	587.395	+ 15.208
9-10 »	539.971	601.732	+ 61.761 (2)

(1) V. *Annali di Statistica*, serie 3a, volume XVI, pag. 11-17, Roma, Bencini, 1885.

(2) Il censimento dei fanciulli da 9 a 10 anni torna a dare una differenza piuttosto rilevante in confronto del calcolo di sopravvivenza, perchè a danno di questa classe si deve aver avuto una agglomerazione in corrispondenza dell'età rotonda, 10 anni, appartenente alla classe da 10 a 11, come lo provano i dati dei capoluoghi di provincia e di circondario, che il censimento del 1881 ha specificati anno per anno di età, anche dopo le prime dieci classi.

Nel calcolo dei sopravviventi, giova avvertirlo, venivano bensì trascurati i movimenti migratorii, ma è risaputo che i bambini vi contribuiscono in assai piccola misura e che d'altronde havvi un certo compenso tra immigrazione ed emigrazione. Sicchè le cifre calcolate in base ad elementi di sicura rilevazione, come le nascite e le morti, erano da ritenersi più attendibili e da proporre come correttivo a quelle del censimento. In occasione del quale forse avviene che buon numero di bambini dati a balia in campagna non siano denunziati nè dai genitori, nè dalla nutrice; che di qualcuno illegittimo, ma allevato in casa, si taccia deliberatamente sulla scheda di famiglia e che si taccia anche di qualcuno legittimo, considerato come *quantité négligeable*.

c) Comparazione fra termini risultanti entrambi, per diverse vie, da calcolo o induzione statistica. — Ad es., l'ammontare probabile della ricchezza fondiaria si può desumere dal contingente d'imposta, adottando un certo moltiplicatore per avere il reddito e un certo saggio di capitalizzazione per avere la cifra capitale; ovvero si può desumere dal valore degli immobili, che cadono annualmente in successione ereditaria, prendendo come moltiplicatore il numero degli anni di durata di una generazione nella classe dei proprietari, o l'intervallo medio tra due successivi trasferimenti *mortis causa*. Se i risultati discordano, rimane a decidere qual è il più attendibile e, supposto risoluto il dubbio, bisogna vedere se nel meno attendibile l'errore sta nell'elemento statisticamente rilevato o nel moltiplicatore, che si è voluto adottare. Ma qui, evidentemente, usciamo dal campo proprio della rilevazione e dei suoi errori, per entrare in quello in cui si discute della bontà relativa di due induzioni.

§ 7. Localizzare, individuare la sede degli errori sospettabili in un quadro di rilevazione, scendendo alla comparazione per classi e per gruppi, misurarne la grandezza di effetto, anche in maniera approssimativa, è l'ufficio della critica in questo stadio. Tale ufficio viene di molto agevolato dai procedimenti semplificativi e rappresentativi delle quantità rilevate; ma non si può dire che questi siano assolutamente indispensabili a quello. Chè anzi la critica preventiva risparmia spesso un lavoro, altrimenti inutile, di elaborazione. I procedimenti, cui si allude, appaiono piuttosto necessarii per un ulteriore còmpito della critica, in uno stadio in cui la rilevazione fa stato, cioè ha il valore di un fatto compiuto, definitivo. Allora sarà il caso di svolgere alcuni punti della teoria matematica degli errori, la quale, presupponendo nozioni speciali intorno alle medie, ai rapporti, alle interpolazioni e al calcolo di probabilità, si troverebbe a disagio in questa sede.

CAPO TERZO

Procedimenti semplificativi e rappresentativi delle quantità rilevate.

TITOLO I.

PROCEDIMENTI ARITMETICI.

A) *Teoria delle medie.*

Sommario: § 1. Specie di procedimenti semplificativi e rappresentativi. — § 2. Il concetto di *media*. — § 3. La media aritmetica e le sue proprietà matematiche. — § 4 e 5. Media geometrica e armonica. — § 6. Competenza d'applicazione delle diverse medie: *A*) nel caso di riduzione di una serie qualunque a serie statica; *B*) nel caso di serie correlate; *C*) nel caso di determinazione del valor più probabile di osservazioni discordanti. — § 7. Osservazioni immediate, mediate e condizionate. — § 8. Scostamento medio e precisione delle serie. — § 9. Medie semplici e ponderate, tipiche e non tipiche, ecc.

§ **1.** Compiuta la rivelazione, esercitata una critica preliminare, si passa alla elaborazione dei dati. Le quantità statistiche greggie, in forma cioè di numeri assoluti, non si prestano bene alla comprensione immediata e chiara dei rapporti di grandezza che esse hanno con altre quantità statistiche e alla scoperta delle uniformità empiriche e delle leggi approssimative; occorre un lavoro preparatorio di semplificazione e di rappresentazione, che ponga nella maggior evidenza possibile tali rapporti, lo stato o l'andamento normale dei fenomeni, la concomitanza o indipendenza delle loro variazioni, ecc. I procedimenti, che fanno al caso, sono di tre specie: *aritmetici*, *geometrici* o *grafici* e *algebrici*.

Cominciamo dai primi, che comprendono le « medie », i « rapporti » e le « perequazioni ». Anzitutto, dunque, delle *medie*.

§ **2.** L'idea di « media » involge generalmente quella di un conguaglio o livellamento di quantità diseguali, ma omogenee. Essa deriva da un bisogno della nostra mente, incapace di ritenere una serie esatta di impressioni riferentisi a diverse grandezze numeriche. Colla media si uniformano, in certa guisa, coteste impressioni; si sostituisce alla serie un termine unico, alla cui formazione concorrono e collaborano tutti i termini dati; qualche cosa insomma di analogo a ciò che è il centro di gravità in un sistema di punti pesanti.

La teoria delle medie ha in Statistica una importanza meno grande di quella che le si assegnava una volta. Perchè l'essenziale nelle ricerche concernenti i fenomeni collettivi è di arrivare alla *legge*, la quale suppone una successione di termini, sia pure sintetizzata in un enunciato breve quanto si vuole. Invece la media, termine unico, occulta la legge della serie (1) o tutt'al più rappresenta la condizione statica del fenomeno. Poi, anche prescindendo da questo, la teoria delle medie presenta parecchi lati oscuri. Illimitato è il numero delle medie e controversa la competenza rispettiva di applicazione; incerto il significato, massime se la media vien calcolata fra termini molto disuguali; arbitrario il numero dei termini sui quali si esercita il calcolo. Ad ogni modo il bisogno di semplificare è così sentito nelle osservazioni dei fenomeni di massa, che i servigi resi dai valori medii debbono essere convenientemente apprezzati.

Tra due o più numeri si possono dare infiniti valori medii, ma quelli che ricorrono più spesso nell'uso pratico sono: la ***media aritmetica***, la ***geometrica*** e l'***armonica***, conosciute già dagli antichi, che le attribuivano a Pitagora. L'algebra, colla generalità dei suoi simboli, dimostra che non v'è limite assegnabile al numero delle medie. Per esempio, chiamando a_1, a_2, a_3..... a_n i termini di una serie, il cui numero è n, l'espressione:

$$\sqrt[m]{\frac{a_1^m + a_2^m + a_3^m + \dots + a_n^m}{n}}$$

dà luogo a tante medie diverse l'una dall'altra, quanti i valori che possiamo far assumere ad m, cioè infiniti. Nel caso particolare in cui si abbia $m = 2$, l'espressione

$$\sqrt{\frac{a_1^2 + a_2^2 + a_3^2 \dots + a_n^2}{n}}$$

fornisce una media molto usata nel calcolo degli errori di osservazione.

§ **3.** La forma più semplice e di pratica comunissima è la *media aritmetica*. Essa si ottiene dividendo la somma dei termini dati per il loro numero. La sua espressione generale per n termini è dunque:

$$\text{Media aritmetica} = \frac{a_1 + a_2 + a_3 + \dots + a_n}{n}.$$

Quindi, ad es., fra i due termini 14 e 18 la media aritmetica sarà:

$$\frac{14 + 18}{2} = 16;$$

(1) A. Messedaglia, *Il calcolo dei valori medi* (*Archivio di Statistica* anno 5°, fasc. 4°, pag. 513, Roma-Torino, E. Loescher, 1880).

fra i termini 22, 28, 31, sarà:

$$\frac{22 + 28 + 31}{3} = 27;$$

fra quattro termini, come 100, 115, 99, 130, sarà:

$$\frac{110 + 115 + 99 + 130}{4} = 113{,}5, \text{ ecc.}$$

Per eseguire più rapidamente il calcolo, quando si tratti di lunghe serie a termini poco diseguali, convien sottrarre da ciascun termine una cifra costante e rotonda e calcolare sulla nuova serie (così ridotta a numeri più piccoli) la media aritmetica, la quale poi, aggiunta alla costante, dà per risultato la media aritmetica della serie primitiva. Ad esempio, dalla serie già riferita 110, 115, 99, 130, sottraendo la costante 100, si ottengono i termini: 10, 15, —1, 30, la cui somma 54, divisa per 4, dà per quoziente 13,5. Aggiungendo 13,5 alla costante 100, si riproduce la media dianzi trovata, cioè 113,5. Questo artificio di semplificazione viene anche meglio in acconcio nel caso di medie ponderate (1).

Le proprietà matematiche, per le quali si segnala la media aritmetica, sono soprattutto le seguenti:

1° *la somma degli scostamenti positivi dei singoli termini dalla media aritmetica eguaglia la somma degli scostamenti negativi;* ossia *la somma algebrica di tutti gli scostamenti è eguale a zero;*

2° *la somma dei quadrati degli scostamenti dei singoli termini dalla loro media aritmetica è un* « minimum » *in confronto della somma di quadrati di scostamenti, a cui darebbe luogo un altro valore qualsiasi diverso dalla media aritmetica.*

Sia, ad esempio, la serie di sette termini: 25, 32, 28, 26, 19, 33, 40, la cui media aritmetica è 29. Gli scostamenti semplici e i quadrati degli scostamenti dànno, come si legge nel prospetto qui a fianco, la somma *zero* e la somma 272, rispettivamente. Si avverta che la elevazione a quadrato fa diventare positivi tutti i termini della terza colonna.

	Serie	Scostamenti positivi (+) e negativi (—)	Quadrati degli scostamenti
	25	— 4	16
	32	+ 3	9
	28	— 1	1
	26	— 3	9
	19	— 10	100
	33	+ 4	16
	40	+ 11	121
Media	29	*Somma* 0	*Somma* 272

(1) Veggasi R. Livi, *Antropometria militare*, pag. 23-25, nota.

Or bene, se avessimo preso un valore qualsiasi diverso da 29, che è la media aritmetica, la somma algebrica degli scostamenti semplici dei singoli termini della serie da cotesto nuovo valore sarebbe riuscita diversa da *zero* e la somma dei quadrati degli scostamenti stessi avrebbe dato un risultato *superiore a* 272, come ciascuno può facilmente verificare.

Ad ogni modo diamo in nota la dimostrazione generale delle due proprietà caratteristiche in parola (1).

In grazia di queste proprietà, la media aritmetica offre, come ora vedremo, un buon criterio di scelta nel caso di osservazioni discordanti, ma meritevoli di eguale fiducia e inoltre serve a ben rappresentare lo stato quantitativo normale del fenomeno nel caso di serie statiche. La facilità e comodità del calcolo, sebbene costituiscano un vantaggio semplicemente subbiettivo del calcolatore, hanno finito per sanzionare l'uso di questa media anche in casi nei quali, a rigore, la competenza di applicazione spetterebbe a medie di diversa natura.

---

(1) Siano $a_1, a_2, a_3 \dots a_n$ i termini di una serie statistica e si chiami M la loro media aritmetica. Si avrà per definizione:

$$M = \frac{a_1 + a_2 + a_3 \dots + a_n}{n}$$

Facciamo gli scostamenti dei singoli termini della media M, chiamando x_1 lo scostamento del primo termine a_1, x_2 quello del secondo, ecc., cioè:

$$\begin{aligned} x_1 &= a_1 - M \\ x_2 &= a_2 - M \\ x_3 &= a_3 - M \\ \vdots\ & \quad \vdots \quad\ \vdots \\ x_n &= a_n - M \end{aligned}$$

Ora, sommando gli scostamenti, avremo:

$$x_1 + x_2 + x_3 + \dots + x_n = a_1 + a_2 + a_3 + \dots + a_n - n\ M$$

ossia sostituendo ad M il suo valore:

$$x_1 + x_2 + x_3 + \dots + x_n = a_1 + a_2 + a_3 + \dots + a_n - $$
$$- n \frac{a_1 + a_2 + a_3 + \dots + a_n}{n} =$$
$$= a_1 + a_2 + a_3 + \dots + a_n - (a_1 + a_2 + a_3 + \dots + a_n)$$

donde evidentemente:

$$x_1 + x_2 + x_3 + \dots + x_n = \textit{zero}.$$

Resta così dimostrata la prima proprietà della media aritmetica, cioè che la somma algebrica degli scostamenti è zero. Passiamo alla seconda:

Siano ancora $x_1, x_2, x_3 \dots x_n$ gli scostamenti dei singoli termini della serie dalla loro media aritmetica; e siano $y_1, y_2, y_3 \dots y_n$ gli scostamenti degli stessi termini da un altro valore, comunque scelto, medio o non medio per rispetto alla serie data. È cosa ovvia che i nuovi scostamenti differiranno dai precedenti di tanto quant'è la differenza tra la media aritmetica

§ 4. L'espressione generale della *media geometrica*, per n termini, è:

$$M_g = \sqrt[n]{a_1 \times a_2 \times a_3 \times \ldots \times a_n}$$

ossia la radice *ennesima* del prodotto degli n termini.

Fra 9 e 12 la media geometrica sarà dunque:

$$\sqrt{9 \times 12} = \sqrt[2]{108} = 10{,}3923\ldots;$$

fra i numeri 7, 10 e 14, sarà:

$$\sqrt{7 \times 10 \times 14} = \sqrt[3]{980} = 9{,}93288\ldots \text{ecc.}$$

Le proprietà, per le quali si distingue la media geometrica e che ne determinano la competenza di applicazione, sono specialmente queste:

1° ***Il quadrato della media geometrica coincide colla media geometrica dei quadrati dei termini;*** e lo stesso dicasi dei cubi, delle quarte

---

e il nuovo valore adottato, cioè differiranno di una quantità costante, che chiameremo z. Si avrà dunque:

$$\begin{aligned} y_1 &= x_1 + z \\ y_2 &= x_2 + z \\ y_3 &= x_3 + z \\ &\vdots \\ y_n &= x_n + z \end{aligned}$$

e z sarà una quantità positiva, se il nuovo valore adottato è inferiore all media aritmetica, negativa nel caso contrario. Ora, elevando a quadrato tali eguaglianze e sommando, avremo:

$$\begin{aligned} y_1^2 &= x_1^2 + z^2 + 2x_1 z \\ y_2^2 &= x_2^2 + z^2 + 2x_2 z \\ y_3^2 &= x_3^2 + z^2 + 2x_3 z \\ &\vdots \\ y_n^2 &= x_n^2 + z^2 + 2x_n z \\ \hline \Sigma(y^2) &= \Sigma(x^2) + nz^2 + 2z\,\Sigma(x). \end{aligned}$$

Ma, per quel che si è dimostrato nella prima parte di questa nota, l'espressione $\Sigma(x)$, cioè la somma degli scostamenti semplici dei singoli termini dalla media aritmetica è eguale a *zero;* quindi sarà *zero* il prodotto $2z\,\Sigma(x)$ e la eguaglianza si riduce a questa forma:

$$\Sigma(y^2) = \Sigma(x^2) + nz^2$$

Si noti che se $\Sigma(x)$ è eguale a zero, non lo è invece $\Sigma(x^2)$, ossia la somma dei quadrati degli scostamenti, perchè colla elevazione a quadrato anche gli scostamenti negativi diventano positivi. Similmente z^2 è una quantità positiva, anche se z è negativa.

L'essere $\Sigma(y^2) = \Sigma(x^2) + nz^2$, ossia maggiore di $\Sigma(x^2)$, significa appunto che la somma dei quadrati degli scostamenti dei singoli termini da un valore qualunque diverso dalla media aritmetica, è sempre maggiore di quella che si ottiene prendendo invece la media aritmetica. Questa sola realizza dunque la condizione del *minimum*, come dovevasi dimostrare.

potenze, ecc., oppure delle radici quadrate, cubiche, ecc. Ciò non si verifica invece per la media aritmetica; ad es.: il quadrato della media aritmetica *non* coincide colla media aritmetica dei quadrati dei termini (1).

2° *Il reciproco della media geometrica coincide colla media geometrica dei reciproci.* — Tale corrispondenza manca invece se si mettono a confronto il reciproco della media aritmetica e la media aritmetica dei reciproci (2).

Gli è per queste proprietà che taluni assegnano alla media geometrica una competenza più estesa che alla media aritmetica.

È facile vedere che se

$$M_g = \sqrt[n]{a_1 \times a_2 \times \ldots \times a_n}$$

prendendo i logaritmi si avrà:

$$\log M_g = \frac{\log a_1 + \log a^2 + \ldots + \log a_n}{n}$$

ossia: *il logaritmo della media geometrica è eguale alla media aritmetica dei logaritmi dei termini.*

---

(1) Infatti $(\sqrt{a\,b})^2$ è manifestamente eguale a $\sqrt{a^2 b^2}$, ossia tra il quadrato della media geometrica e la media geometrica dei quadrati vi è identità. Del pari abbiamo

$$(\sqrt{a\,b})^3 = \sqrt{a^3 b^3}$$

e in generale poi

$$(\sqrt{a\,b})^m = \sqrt{a^m b^m}.$$

Invece $\left(\frac{a+b}{2}\right)^2$, cioè il quadrato della media aritmetica, non può essere eguale ad $\frac{a^2+b^2}{2}$ (media aritmetica dei quadrati dei termini), salvo nel caso in cui sia $a = b$, come sarebbe facile dimostrare.

(2) Dati due termini a e b, la loro media geometrica è $\sqrt{a\,b}$; il reciproco di tal media è $\frac{1}{\sqrt{ab}}$. Facciansi ora i reciproci dei termini dati, cioè $\frac{1}{a}$ e $\frac{1}{b}$; la media geometrica dei reciproci sarà $\sqrt{\frac{1}{a}\cdot\frac{1}{b}}$. Ed evidentemente sussisterà l'eguaglianza:

$$\frac{1}{\sqrt{a\,b}} = \sqrt{\frac{1}{a}\cdot\frac{1}{b}}.$$

Al contrario il reciproco della media aritmetica, espresso da $\frac{1}{\left(\frac{a+b}{2}\right)}$, non è eguale alla media aritmetica dei reciproci, che è $\frac{\frac{1}{a}+\frac{1}{b}}{2}$, salvo il caso in cui sia $a = b$, come con facili operazioni si può verificare.

§ 5. La media armonica fra n termini ha per espressione generale:

$$M_{\text{arm.}} = \frac{n}{\frac{1}{a_1} + \frac{1}{a_2} + \ldots + \frac{1}{a_n}}$$

e potrebbe definirsi: *il reciproco della media aritmetica dei reciproci.* Infatti l'espressione data deriva da quest'altra:

$$\frac{1}{\left(\frac{\frac{1}{a_1} + \frac{1}{a_2} + \ldots + \frac{1}{a_n}}{n}\right)}$$

in cui al numeratore figura l'unità e al denominatore la media aritmetica dei reciproci dei termini.

Nel caso di due termini a e b, la media armonica è dunque:

$$\frac{2}{\frac{1}{a} + \frac{1}{b}} = \frac{2ab}{a + b},$$

cioè il doppio prodotto diviso per la somma. Ad esempio, il cubo ha 6 faccie, 8 angoli e 12 lati; i numeri 6, 8, 12 diconsi in proporzione armonica, perchè il termine di mezzo, 8, è appunto medio armonico tra 6 e 12. Infatti $8 = \frac{2 \times 6 \times 12}{6 + 12}$.

§ 6. Queste le tre medie classiche, che gli scrittori d'altri tempi elevarono persino all'onore di simboli delle tre forme di governo: democrazia, aristocrazia e monarchia temperata. Il Messedaglia cita in proposito Severino Boezio (*De institutione arithmetica*) e Giovanni Bodin (*De republica*) e ricorda come anche il grande Keplero nell'opera sull'*Armonia del mondo* non isdegnasse dedicare, a sollievo dei lettori, una lunga digressione politica sulle tre medie e ricercare le applicazioni dell'armonica alle cose sensibili e alle intellettuali e morali (1). Ma noi dobbiamo qui dire della competenza statistica delle medie, competenza determinata dalle proprietà loro matematiche e dal genere di relazioni che intercedono tra i fenomeni in esame.

Ora, per ben chiarire la questione di competenza, bisogna distinguere il caso in cui si voglia ridurre una serie qualunque di dati statistici ad una serie teorica regolare, d'egual somma e d'egual numero di termini, dal caso in cui si considerino due serie collegate ossia

---

(1) Ci sarebbe da parlare pure della *media antiarmonica*, la quale per due termini è eguale alla somma dei quadrati divisa per la somma dei termini semplici: M. antiarm. $= \frac{a^2 + b^2}{a + b}$. Ma per questi ed altri particolari rinviasi il lettore alla citata monografia del Messedaglia.

vincolate da un qualsiasi rapporto di dipendenza, e infine dal caso in cui si tratti di ricavare il valore più probabile dei dati in un sistema di osservazioni discordanti riferibili ad uno stesso oggetto.

A) Allorquando una serie statistica, che figura il movimento di un fenomeno collettivo nel tempo, si vuol ridurre a ***serie perfettamente statica***, d'egual somma e d'egual numero di termini o, in altre parole, si vuol livellare in guisa che i diversi termini diventino tutti di pari grandezza, la competenza della media aritmetica è assoluta. Perocchè esiste un solo modo di operare quella trasformazione e il modo sta appunto nel sostituire ai singoli dati dell'osservazione la media aritmetica loro. La quale rappresenta così una ***costante non arbitraria***, chè fornisce non solo il valor centrale della serie teorica, ossia quello che corrisponde alla metà precisa dell'intervallo, ma insieme i valori corrispondenti agli altri punti di divisione dell'intervallo.

Se, invece, la serie statistica data si vuol ridurre a ***serie ancora regolare, ma dinamica***, avverrà bensì che la media geometrica o l'armonica o l'antiarmonica od altra non specificamente denominata e perfino la stessa media aritmetica (nelle serie *lineari*) fornisca il valor centrale, ma non già che fornisca al tempo stesso i restanti valori della serie teorica, i quali debbono generalmente differire l'uno dall'altro e dal valor centrale (1); altrimenti la serie non sarebbe più dinamica. È questa una considerazione necessaria nella soggetta questione.

Inoltre, mentre di serie statiche d'egual somma ed egual numero di termini della serie osservata ce n'è una sola, sì che non esiste arbitrio di scelta, di serie dinamiche pure d'egual somma ed egual

---

(1) Nelle serie dinamiche dette *lineari*, costituenti cioè una progressione aritmetica (come, ad esempio, i numeri 12, 15, 18, 21, 24 ...), la media aritmetica dà il valor centrale o corrispondente alla metà precisa dell'intervallo, ma non dà più, come nelle serie statiche, i restanti valori. Non c'è nemmeno bisogno di dimostrarlo. Così, nelle serie *geometriche*, costituenti cioè una progressione geometrica (come i numeri 2, 6, 18, 54 ...), il termine centrale è dato dalla media geometrica, e s'intende, il termine centrale soltanto. In queste poi la media aritmetica corrisponderebbe ad un punto non più centrale della serie. Sia, ad esempio, la progressione: a, aq, aq^2, aq^3, aq^4. La media geometrica, aq^2, è precisamente il valore della metà intervallo; invece la media aritmetica, la quale nel caso concreto avrebbe per sua espressione:

$$\frac{a}{5}\left(\frac{1-q^5}{1-q}\right),$$

dà un valore maggiore e quindi corrispondente ad un punto situato più innanzi o più addietro, secondo i casi, nella serie dei tempi.

Supponiamo ancora una serie parabolica di 2° grado, in cui cioè le differenze prime dei termini costituiscono una progressione aritmetica e le differenze seconde (cioè le differenze delle differenze) sono una quantità costante. Se si tratta, ad esempio, di una serie di 5 termini, questi, come meglio vedremo

numero di termini della serie osservata, che a questa si possono sostituire, ce n'è una varietà infinita; donde l'arbitrio della scelta, il quale arbitrio (come meglio vedremo nel capitolo dell'interpolazione) viene disciplinato, ma non eliminato, quando noi ci vincoliamo ad una condizione speciale, per es.: alla condizione che la somma dei quadrati degli scostamenti fra i termini della serie osservata e quelli della serie teorica sostituita, sia un *minimum*. All'opposto la competenza della media aritmetica per le serie statiche non deriva dal fatto che essa realizza già per se stessa la condizione dei minimi quadrati in confronto d'ogni altra costante (secondo quanto fu dimostrato nella nota a pag. 94-95), ma dal fatto che *una sola* è la serie perfettamente statica, d'egual somma e d'egual numero di termini, sostituibile alla serie osservata.

In realtà, qualunque serie statistica, per piccole oscillazioni che presenti da una ad altra unità di tempo in cui è diviso l'intervallo, ha in ultima analisi carattere dinamico; ma essa può sempre concepirsi come la risultante di più forze componenti, di cui una agisca in modo uniforme nel tempo e l'altra o le altre in maniera più o meno complessa. Così la serie:

25, 17, 12, 10, 11, 15, 22

può considerarsi come risultante dalla composizione di queste due:

16, 16, 16, 16, 16, 16, 16
9, 1, —4, —6, —5, —1, 6

di cui la prima è una serie costante, a termini eguali tra loro ed eguali alla media aritmetica dei dati originarii; la seconda è una serie parabolica di somma *zero*, espressione di cause che possono agire ora positivamente, ora anche negativamente, cioè in controsenso di quelle che avrebbero prodotto la serie costante. Ebbene l'ipotesi, che ha presieduto a una siffatta decomposizione, può essere irreale o verosimile in svariatissimi gradi, secondo i casi; ma è sempre ipotesi lecita, purchè da essa non si pretenda di ricavare conseguenze maggiori di

---

a suo luogo, possono algebricamente esprimersi coll'equazione: $ax^2 + bx + c$, in cui la variabile x assume successivamente i valori — 2, — 1, 0, + 1, + 2. Ne risultano le espressioni:

$$\begin{array}{r} 4a - 2b + c \\ a - b + c \\ c \\ a + b + c \\ 4a + 2b + c \\ \hline \end{array}$$

$$\begin{array}{ll} \textit{Somma} \ \ldots & = 10a + 5c \\ \textit{Media aritmet.} & = 2a + c \end{array}$$

Il valor centrale della serie è c, mentre la media aritmetica dei termini, che è $= 2a + c$, differisce dal valor centrale e corrisponde necessariamente ad altro punto dell'intervallo.

quelle che comporta. Evidentemente essa deve allontanarsi tanto meno dalla realtà, quanto più piccole sono le differenze fra un termine e l'altro della serie osservata e quanto più saltuariamente positive e negative (1). In tal caso la costante non arbitraria fornita dalla media aritmetica esprime bene lo *stato quantitativo normale* del fenomeno, mentre le residue oscillazioni rivestono il carattere di semplici perturbazioni dovute a cause secondarie.

Prendiamo la serie delle legittimazioni di figli naturali avvenute in Italia nel novennio 1893-1901. Essa può decomporsi nella serie costante di termini eguali alla media aritmetica 23.826,2 e nella serie degli scostamenti.

ANNI	Serie osservata	COMPONENTI	
		Serie statica	Serie degli scostamenti
1893	22.683	23.826,2	— 1.143,2
1894	22.913	23.826,2	— 913,2
1895	24.282	23.826,2	+ 455,8
1896	25.387	23.826,2	+ 1.560,8
1897	23.596	23.826,2	— 230,2
1898	22.659	23.826,2	— 1,167,2
1899	24.686	23.826,2	+ 859,8
1900	23.457	23.826,2	— 369,2
1901	24.773	23.826,2	+ 946,8
Somma	214.436	214.436,0	—
Media aritmetica	23.826,2	—	—

Questi scostamenti non sono rilevanti; la loro media grandezza, calcolata facendo astrazione dal segno, non raggiunge il 4 % della media delle quantità osservate; e il maggiore tra essi, in via assoluta e relativa, rappresenta appena il 6 % del corrispondente termine. Inoltre gli scostamenti di segno positivo e quelli di segno negativo si succedono in modo piuttosto saltuario, come a caso. Ne inferiamo pertanto che la media aritmetica rappresenta qui abbastanza bene la condizione quantitativa normale del fenomeno, ben inteso nei limiti di tempo considerati, e che le oscillazioni della serie hanno carattere di perturbazioni prodotte da cause accidentali, secondarie, intermittenti.

Riassumendo: quando si adotta una media determinata, a preferenza di altre, come valore corrispondente alla precisa metà dell'in-

(1) Dicasi pure, anche quando le variazioni presentano un deciso carattere di periodicità

tervallo, è implicito che si sia anticipatamente scelta la serie teorica, di cui la media in parola costituisce il termine centrale. Quindi, a rigore, una trattazione compiuta delle medie dovrebbe susseguire a quella delle serie. D'altra parte l'arbitrio di scelta della serie teorica dev'essere disciplinato in qualche modo da un attento esame della forma della serie in osservazione, dalle conoscenze che anche per altra via possiamo avere della dinamica del fenomeno, ecc.

Suppongasi di voler determinare il numero probabile di abitanti d'un paese a metà intervallo tra due censimenti. Le ipotesi, che si possono fare, sono molte, ma le più ovvie e semplici sono queste due: 1° che la popolazione sia cresciuta (tralasciamo il caso di diminuzione) tra un censimento e l'altro di quantità assolute eguali ad ogni unità di tempo, per esempio, l'anno; 2° o che la popolazione sia cresciuta di quantità regolarmente crescenti d'anno in anno. Accorderemo la preferenza alla prima ipotesi, se, poniamo, le statistiche del movimento dello stato civile ci informano che l'eccedenza delle nascite sulle morti è stata piuttosto costante nell'intervallo e se non vi sia ragione di credere che tale costanza abbiano perturbata i movimenti migratorii; allora la popolazione alla metà del periodo sarà data dalla semi-somma delle cifre dei due censimenti, cioè dalla loro media aritmetica. La seconda ipotesi dovrà invece esser preferita in caso che l'eccedenza delle nascite sulle morti risulti crescente con qualche regolarità d'anno in anno; la qual cosa significherebbe che la popolazione tende a svilupparsi nella maniera stessa di un capitale investito, non ad interesse semplice, ma ad interesse composto. Come gli interessi, capitalizzandosi alla fine d'ogni unità di tempo, diventano produttivi di nuovi interessi, così l'eccedenza continuata delle nascite sulle morti riuscendo ad incremento di tutte le classi d'età, e quindi anche di quelle atte alla procreazione, diviene produttiva di ulteriori eccedenze. Ora, siccome un capitale ad interesse composto si svolge in progressione geometrica secondo la nota formola: $C_n = C(1+t)^n$, questa formola potrà applicarsi alla popolazione (1). La media geome-

---

(1) Ricordiamo per comodità del lettore la derivazione di questa formola. Sia t l'interesse nell'unità di tempo per una unità di capitale. Per C unità di capitale, in una unità di tempo, l'interesse sarà Ct. Quindi al compimento della prima unità di tempo il capitale originario accresciuto dell'interesse sarà diventato $C + Ct = C(1+t)$.

Ora l'interesse della seconda unità di tempo dovrà calcolarsi sul nuovo capitale $C(1+t)$: esso in altri termini sarà $= t \cdot C(1+t)$, e al compimento della seconda unità il capitale che alla fine della prima era $C(1+t)$, diventerà:

$$C(1+t) + t \cdot C(1+t) = C(1+t)^2.$$

Ripetendo lo stesso ragionamento per n, unità di tempo, si troverebbe:

$$C_n = C(1+t)^n.$$

Questa formola serve a calcolare il capitale finale, quando sia dato l'iniziale, il saggio di interesse e il numero delle unità di tempo; o il capitale

trica tra le cifre dei due censimenti darà la popolazione alla metà intervallo. Se l'intervallo è, per esempio, di 10 anni, chiamando P la popolazione iniziale, si avranno i tre termini:

$$P, \quad P(1+t)^5, \quad P(1+t)^{10}$$

di cui il secondo, rappresentante la popolazione alla fine del 5° anno (cioè alla metà esatta del decennio), è manifestamente medio geometrico fra il primo e il terzo.

Facciamo un altro caso. La popolazione italiana censita al 31 dicembre 1871 era di 26.801 mila abitanti; quella censita al 31 dicembre 1881 di 28.460 mila e quella calcolata al 31 dicembre 1901, in base al censimento del 9-10 febbraio di quello stesso anno, di 32.689 mila. Si domanda: quanta dev'essere stata la popolazione al 31 dicembre 1886, cioè al punto di mezzo del trentennio 1872-1901? Poichè si hanno tre dati, utilizziamoli tutti tre facendoli concorrere alla determinazione probabile del valor centrale; questo è un modo logico di operare, dal momento che la popolazione del 1901 era in grande parte la medesima del 1881 e in parte ancor notevole quella stessa del 1871. Altra ipotesi ragionevole, che possiamo fare, è quella di un movimento continuo, variante per gradi e non per bruschi salti. Ora, anticipando su ciò che meglio sarà chiarito nel capitolo dell'interpolazione, il metodo più indicato è di far passare una parabola per gli estremi delle tre ordinate che su un diagramma ordinario rappresenterebbero le cifre della popolazione alla fine degli anni 1871, 1881 e 1901. Cioè basta considerare la popolazione come funzione del tempo e porre i risultati dei censimenti sotto forma di sviluppo della funzione $ax^2 + bx + c$, tenuto conto delle ineguali distanze di tempo.

Facendo quindi assumere alla x i valori -3, -2, -1, 0, $+1$, $+2$, $+3$ corrispondenti ad intervalli regolari di 5 anni ciascuno, otteniamo le espressioni:

Date		Popolazione	Migliaia di abitanti
31 dicembre	1871	$9a - 3b + c =$	26.801
» »	1876	$4a - 2b + c$	
» »	1881	$a - b + c =$	28.460
» »	1886	c	
» »	1891	$a + b + c$	
» »	1896	$4a + 2b + c$	
» »	1901	$9a + 3b + c =$	32.689

---

iniziale, quando sia dato il finale e sian dati gli altri due elementi; e così ancora, il saggio d'interesse o il numero delle unità di tempo quando siano dati gli altri elementi. Infatti si ha:

$$C_n = C\,(1+t)^n; \quad C = \frac{C_n}{(1+t)^n}; \quad t = \sqrt[n]{\frac{C_n}{C}} - 1;$$

$$n = \frac{\log \frac{C_n}{C}}{\log (1+t)}.$$

Di esse riproduciamo qui quelle utili pel calcolo:

$$9a - 3b + c = 26.801$$
$$a - b + c = 28.460$$
$$9a + 3b + c = 32.689$$

Risolvendo questo semplice sistema di tre equazioni con tre incognite, si ottiene: $a = 37{,}_{96}$; $b = 981{,}_{33}$; $c = 29.403{,}_{37}$. E siccome la popolazione alla fine del 1886 ha per espressione algebrica c, il suo valore numerico sarà 29.403 mila abitanti. In simil modo, applicando i valori di a, b, e c alle altre formole, si troverebbe la popolazione alla fine degli anni 1876, 1891 e 1896.

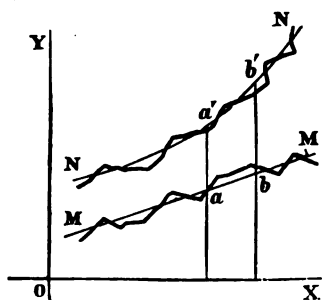
La cifra di 29.403 mila abitanti per il momento centrale del trentennio è senza dubbio la più attendibile; essa differisce, sebbene di poco, dalla media aritmetica dei tre censimenti (= 29.317) e dalla media geometrica (= 29.214). E qui conviene avvertire (punto rilevante nella teoria delle medie) che la media aritmetica e geometrica non si potrebbero applicare *sic et simpliciter* ai risultati dei tre censimenti, trovandosi questi a distanze di tempo ineguali l'uno dall'altro, perchè il secondo è venuto un decennio dopo il primo e il terzo invece un ventennio dopo il secondo. Il modo corretto di fare, ad esempio, la media aritmetica sarebbe di sommare le 7 espressioni algebriche indicate a pagina precedente e dividere la somma per 7, il che darebbe:

$$\text{Media aritmetica} = \frac{28a + 7c}{7} = 4a + c$$

e applicando i valori trovati per a e per c, si otterrebbe: media aritmetica = 29.555 migliaia, il qual valore, ben inteso, non corrisponderebbe più al punto centrale della serie parabolica.

Si potrebbe complicare anche più il problema e immaginare 4 censimenti a distanze non eguali l'uno dall'altro; e allora, considerando ancora la popolazione come funzione del tempo, si farebbe passare per i 4 punti dati una curva di 3° grado ($y = ax^3 + bx^2 + cx + d$), di cui facilmente si determinerebbe il valor centrale. Ma per il lettore non esercitato in questi calcoli la cosa riuscirà più chiara dopo lo svolgimento della materia delle interpolazioni.

B) La seconda questione che interessa nella teoria dei valori medii riguarda il caso di serie collegate, ossia vincolate l'una all'altra da un qualsiasi rapporto di dipendenza o interdipendenza. Il solo principio generale che in proposito si possa formolare è forse questo: « *Se si sceglie un determinato valor medio nella prima serie, non è più arbitraria la scelta del valore corrispondente nelle serie collegate con quella per qualsivoglia forma di funzione* ». Nel diagramma a pagina seguente s'immagini figurato il fenomeno M, che, astrazion fatta da piccole oscillazioni, mostra di seguire un andamento rettilineo, e il fenomeno N

che, pur astrazion fatta da piccole oscillazioni, si svolge con un andamento curvilineo. Se correlazione o interdipendenza esiste fra

d N, è chiaro che, scelto il valor medio *a* nella prima serie, il corrispondente valor medio nella seconda sarà a^1 e non altro: e se invece tra gli infiniti valori medii della prima scegliamo il punto *b*, non sarà più arbitraria la scelta del corrispondente sulla curva N; questo valore sarà b^1 e non altro.

Ora interessa vedere qualche caso speciale di corrispondenza di medie. Uno ce lo porgono le serie riguardanti fenomeni, che si comportano in ragione ***inversamente proporzionale*** gli uni degli altri. Alla media aritmetica dei primi corrisponde allora non la media aritmetica dei secondi, ma la media armonica. Togliamo dal Messedaglia qualche esempio:

Con 10.000 lire si possono comperare 100 unità di merce, quando il prezzo unitario è di 100; ma se il prezzo unitario sale a 150, le diecimila lire non serviranno che a comperare 66 unità e $^2/_3$ (infatti: $66\,^2/_3 \times 150 = 10.000$); e se il prezzo si eleva a 200, non si acquisteranno che 50 unità. Ora delle due serie correlate:

Prezzi	100,	150,	200
Quantità . . .	100,	66 $^2/_3$,	50

la prima è una proporzione *aritmetica*, la seconda è una proporzione *armonica*; ossia in quella il termine di mezzo è medio aritmetico fra gli estremi; in questa è medio armonico (1).

Similmente in un titolo circolante ad interesse nominale *fisso* e a corso di Borsa variabile, l'interesse effettivo è in ragione inversa del corso del titolo. Supponendo che il titolo sia costituito in 5 %, se il prezzo d'acquisto in Borsa scende da 100 a 80, l'interesse effettivo per chi ha acquistato ad 80, non sarà più di 5 %, ma di 6,25 % (infatti $5 : 80 = 6{,}25 : 100$). Se poi il corso scendesse a 60, l'acquirente a questo prezzo verrebbe ad investire il suo capitale all'8,33 %. Quindi alla serie *aritmetica* delle quotazioni (100, 80, 60) corrisponde una serie d'interessi effettivi (5 %, 6,25 %, 8,33 %) che non è più aritmetica, ma *armonica*, come facilmente si può verificare.

È noto, come legge d'acustica, che il numero delle vibrazioni di una corda, in tempi eguali, rimanendo costante la tensione, è inversamente proporzionale alla lunghezza della corda. Così nell'*accordo*

(1) Infatti $66\,^2/_3 = \frac{2 \times 100 \times 50}{100 + 50}$ secondo la formola, data sopra, della media armonica fra due termini.

perfetto maggiore, che si compone di una *terza maggiore* e di una *quinta*, si ha:

	DO	MI	SOL
Lunghezze relative delle corde . .	1	$^4/_5$	$^2/_3$
Numero relativo di vibrazioni . .	1	$^5/_4$	$^3/_2$
Numero assoluto di vibrazioni nella ottava di due piedi	264	330	396

Anche qui è facile verificare che alla media aritmetica dei numeri relativi o assoluti delle vibrazioni ($^5/_4$ è medio aritmetico fra 1 e $^3/_2$; e così 330 lo è fra 264 e 396) corrisponde la media armonica della lunghezza delle corde ($^4/_5$ è medio armonico fra 1 e $^2/_3$).

Sicchè il principio, che altri accoglie senza critica, della competenza generale della media aritmetica trova limitazioni considerevoli in questa prima categoria di serie collegate. In generale anzi, quando le serie sono vincolate tra loro per relazioni matematiche, che non siano la proporzionalità semplice e diretta, la media aritmetica deve avere per corrispondente una media di diversa natura. Così, un circolo, che sia medio aritmetico *per raggio* fra due altri, non può essere contemporaneamente medio aritmetico *per superficie*. Infatti, le superfici stando fra loro come i quadrati dei raggi, si avrà:

Raggi	1, 2, 3
Superfici . . .	1, 4, 9

e si vede subito che 4 *non è* medio aritmetico fra 1 e 9, mentre nella prima serie il 2 *è* medio aritmetico fra 1 e 3.

Similmente l'individuo che sia medio aritmetico per statura in un gruppo di soggetti osservati, non può « in via normale » essere medio aritmetico anche per peso in questo stesso gruppo; perchè i pesi, negli adulti, stanno fra loro come i quadrati delle stature. Nel fatto quest'individuo medio aritmetico così per peso come per statura potrebbe esistere, ma rappresenterebbe non il tipo o la regola, ma un caso di deviazione dalla regola di correlazione dei due caratteri.

Bene dunque il Messedaglia rilevava l'assurdo teorico di supporre l'uomo medio aritmetico ad un tempo per tutte le sue qualità, intercedendo fra queste delle correlazioni speciali diverse dalla proporzionalità semplice e diretta. Ma le incompatibilità matematiche nascenti dall'impiego della media aritmetica sparirebbero, se in luogo di questa si adottasse per le serie correlative la media geometrica. Dicemmo infatti che il quadrato, il cubo, ecc. della media geometrica coincidono colla media geometrica dei quadrati, dei cubi, ecc. dei termini; che il reciproco di tal media coincide colla media geometrica dei reciproci. Quindi il circolo, che è medio geometrico per raggio fra altri due, lo è pure per superficie; la corda, che è media geometrica per lunghezza tra altre due, dà un numero di vibrazioni che è medio geometrico tra quelli delle altre due corde; e, analogamente, coinciderebbero la media geometrica delle stature e quella dei pesi in un

gruppo di individui. Il concetto dell'*uomo medio geometrico* per tutte le sue qualità ad un tempo, non sarebbe più un assurdo dal punto di vista teorico. Quest'è il motivo pel quale alcuni autori riconoscono alla media geometrica una competenza d'applicazione più estesa che alla media aritmetica.

Tuttavia, noi insistiamo, il solo principio veramente generale che si può enunciare in fatto di serie collegate per qualche relazione causale, è che scelto un qualunque valor medio in una di esse, non è più arbitraria la scelta del corrispondente valore nella seconda, ma dipende dalla legge di questa.

C) Il terzo punto riguarda il modo di ricavare da osservazioni discordanti fatte su uno stesso obbietto il valore più probabile.

Supponiamo d'aver misurato venti volte l'altezza di un monte sul livello del mare e d'avere ottenuto venti risultati poco o molto discordi l'uno dall'altro. Ciò vorrà dire che delle cause d'errore si sono insinuate, nonostante ogni precauzione, nei diversi rilievi. Orbene, se non esistono motivi plausibili per preferire il risultato di una certa osservazione, scartando le altre, se insomma tutte le osservazioni, per essere state condotte con egual diligenza, meritano eguale fiducia, la media aritmetica delle venti misure fornirà il miglior criterio di scelta. Ma perchè la media aritmetica e non la geometrica o l'armonica o l'antiarmonica od altra non denominata? Perchè, rispondiamo, la media aritmetica è la sola *costante*, la quale ripetuta per tante volte quante sono le osservazioni fatte, riproduce esattamente la somma dei risultati ottenuti; quindi essa presenta la più stretta analogia di caratteri col valor vero cercato, il quale è senza dubbio una *costante* (1). L'altezza reale del monte è quella che è e non altra. Se le osservazioni fossero riuscite rigorosamente esatte, esse avrebbero dato risultati uniformi, ossia una costante; ma non essendo riuscite tali, *l'unico modo di ridurle a serie costante d'egual somma e d'egual numero di termini è appunto quello di sostituire ad ogni termine dato la media aritmetica di tutti i termini.* Qui adunque si prescinde dal fatto che la media in parola ha la proprietà di render nulla la somma degli scostamenti semplici e minima la somma dei quadrati degli scostamenti, in confronto di qualsiasi altro valore, medio o non medio, assunto come costante; queste proprietà sono importantissime, ma, meglio che criteri per arrivare al principio della media aritmetica, sono delle conseguenze del principio.

Una questione sottile, sollevata già da matematici, è se la media aritmetica si possa accogliere, oltre che come criterio razionale di scelta tra osservazioni discordanti, meritevoli peraltro d'eguale fiducia,

---

(1) Veggasi ANNIBALE FERRERO, *Esposizione del metodo dei minimi quadrati*, pag. 6 e 42, Firenze, Barbèra, 1876.

anche come *valore più probabile dei dati*. Alcuni, come il Bertrand (1), lo contestano. I singoli risultati hanno una probabilità indeterminata, che si ritiene eguale per tutti se le osservazioni furono fatte colla egual cura; quindi una probabilità nel senso soggettivo soltanto. Ora come può un semplice spediente di calcolo farla diventare oggettiva? La probabilità non si accresce se non a condizione che, restando invariati i casi possibili, crescano i casi favorevoli, ovvero, restando invariati i casi favorevoli, diminuiscano i casi possibili; ma l'operazione della media aritmetica non altera la situazione. Questo ci sembra il pensiero dei critici. Tuttavia crediamo dimostrabile la correttezza dell'espressione « valore più probabile dei dati », trattando le diverse osservazioni come se fossero testimonianze egualmente attendibili. Spieghiamoci con un esempio.

Siano i numeri:

87, 83, 84, 93, 86, 81, 101, 87, 90, 93

i risultati di una serie di 10 osservazioni concernenti la misura d'un dato oggetto. La media aritmetica è 88,5. Dico che essa è il valor più probabile della misura cercata. All'uopo dispongo i singoli risultati delle osservazioni in serie secondo l'ordine di grandezza, dal più piccolo al più grande, e ragiono così:

$$\begin{aligned}
81 &= 81\\
83 &= 81 + 2\\
84 &= 81 + 2 + 1\\
86 &= 81 + 2 + 1 + 2\\
87 &= 81 + 2 + 1 + 2 + 1\\
87 &= 81 + 2 + 1 + 2 + 1\\
90 &= 81 + 2 + 1 + 2 + 1 + 3\\
93 &= 81 + 2 + 1 + 2 + 1 + 3 + 3\\
93 &= 81 + 2 + 1 + 2 + 1 + 3 + 3\\
101 &= 81 + 2 + 1 + 2 + 1 + 3 + 3 + 8
\end{aligned}$$

Media 88,5

$$\begin{aligned}
{}^{10}/_{10} \cdot 81 &= 81\\
{}^{9}/_{10} \cdot 2 &= 1{,}8\\
{}^{8}/_{10} \cdot 1 &= 0{,}8\\
{}^{7}/_{10} \cdot 2 &= 1{,}4\\
{}^{6}/_{10} \cdot 1 &= 0{,}6\\
{}^{4}/_{10} \cdot 3 &= 1{,}2\\
{}^{3}/_{10} \cdot 3 &= 0{,}9\\
{}^{1}/_{10} \cdot 8 &= 0{,}8
\end{aligned}$$

Somma 88,5

Che l'81 faccia parte del valor vero cercato, lo affermano tutte 10 le osservazioni, tutte 10 le testimonianze. Quindi questo elemento entrerà in calcolo in tutta la sua integrità. Scriviamo perciò nel prospetto: ${}^{10}/_{10}\ 81 = 81$.

Che 2 altre unità, oltre l'81, facciano parte del valor vero cercato, lo affermano 9 osservazioni su 10; una lo nega. Dunque, se le osservazioni o testimonianze, per l'ipotesi data, sono egualmente attendibili, le 2 unità entreranno in calcolo solo per ${}^{9}/_{10}$ del loro valore, cioè come 1,8.

Che 1 altra unità, in più degli elementi già calcolati, faccia parte del valor vero, è affermato da 8 osservazioni o testimonianze, su 10;

(1) *Calcul des probabilités*, pag. 176 e segg., Parigi, Gauthier Villars, 1889.

due lo negano. Quell'unità non si accoglierà integralmente in calcolo, ma solo per $^8/_{10}$ del suo valore, ossia come 0,8.

Si continui così fino alle 8 unità per cui differiscono le osservazioni penultima ed ultima in ordine di grandezza. Che quelle 8 unità debbano aggiungersi agli elementi precedentemente calcolati, è affermato da *una* sola testimonianza su dieci e negato da nove; esse non entreranno dunque in calcolo che per un decimo del loro valore, ossia come 0,8.

Sommiamo ora i diversi elementi ammessi; otterremo la media aritmetica della serie primitiva: 88,5. La quale è evidentemente, *nei limiti delle osservazioni fatte*, il valore più probabile, perchè in suo confronto ogni singolo dato o manca di qualche cosa che per altre testimonianze (osservazioni) dovrebbe far parte del valor vero, o contiene qualche cosa di troppo, che dovrebb'essere escluso dal conto. Peraltro il dire che la media aritmetica costituisce, nei limiti della esperienza fatta, il valore più probabile, non significa ancora che si tratti di una probabilità determinata; questa determinazione non si può nemmeno concepire, dal momento che la prima osservazione della serie, la prima in ordine di grandezza, da cui prendemmo le mosse per il nostro calcolo, ha essa medesima un grado ignoto di approssimazione al vero.

Il principio ora dimostrato vale naturalmente pei casi in cui gli errori di osservazione si verificano senza regola conosciuta e si considerano in serie autonome. Altrimenti sarebbe, se l'errore potesse supporsi in un particolare rapporto col nostro modo di percepire o col modo di funzionare degli strumenti registratori, ecc. Allora la competenza della media aritmetica incontra i limiti di cui già dicemmo a proposito delle serie correlate o collegate. Se una bilancia non indicasse direttamente il peso dell'oggetto, ma il logaritmo del peso, è chiaro che un errore di lettura in più ed un errore in meno della stessa grandezza e quindi della stessa probabilità avrebbero nel calcolo del peso reale conseguenze diversamente gravi; e se si adotta la media aritmetica di più letture in logaritmi, ciò implica che si escluda la media aritmetica come valor più probabile dei pesi reali corrispondenti. Così, sia dato un angolo di 30° esatti: un osservatore poco pratico del goniometro abbia la stessa probabilità di errare di un grado in più o di un grado in meno, cioè di valutare l'ampiezza dell'angolo in 29° o in 31°. Qui, ad errori eguali e contrari nella lettura dei gradi, corrispondono necessariamente errori diseguali nel calcolo di tutti i rapporti trigonometrici. Prendiamo, ad esempio, il seno dell'angolo; si avrà:

	Differenze
Sen 29° = 0,48481	— 0,01519
Sen 30° = 0,5	
Sen 31° = 0,51504	+ 0,01504

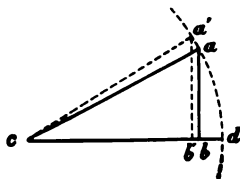
L'errore in meno (— 0,01519) è maggiore dell'errore in più (+ 0,01504), pur avendo la stessa probabilità di verificarsi (1).

Posto che fosse vera la legge di Weber, per cui le sensazioni variano come i logaritmi degli stimoli, ossia le prime si comportano secondo una progressione aritmetica (es.: come 1, 2, 3, 4, 5...) allorchè le cause eccitanti seguono una progressione geometrica (come 1, 2, 4, 8, 16...), evidentemente l'errore di una unità in eccesso, nel giudizio della sensazione ricevuta, avrebbe per i risultati obbiettivi conseguenze più gravi che non l'errore di una unità in difetto.

§ 7. Noi abbiamo trattato qui il caso elementare dell'applicazione della media aritmetica alle osservazioni *immediate*, a quelle cioè in cui l'incognita è unica e fornita direttamente dalle osservazioni, salvo che queste essendo inquinate da errori di ignota natura discordano l'una dall'altra. A suo luogo discorreremo della ricerca dei valori più probabili nelle osservazioni *mediate*, dove le incognite sono due o più e debbono ottenersi da equazioni in numero maggiore delle incognite stesse, i coefficienti essendo forniti da osservazioni poco o molto discordanti. È risaputo che se le equazioni sono in numero minore delle incognite il problema è *indeterminato;* se sono in numero eguale, esso è *determinato* e si risolve coi procedimenti ordinari d'eliminazione; se infine, come nel caso di cui si tratterà, le equazioni sono in numero maggiore delle incognite, il problema si dice *più che determinato.* Vi ha poi una terza categoria di osservazioni dette *condizionate*, nelle quali le quantità empiricamente date debbono soddisfare a tutto rigore a certe condizioni espresse da equazioni; ad esempio: quando, osservati ripetute volte i tre angoli di un triangolo in una operazione geodetica, si deve stabilire la condizione necessaria che la somma dei tre angoli sia eguale esattamente a due retti, ciò che le medie aritmetiche delle misure ripetute sui singoli angoli potrebbero non dare esattamente.

§ 8. La proprietà della media aritmetica, di render minima la somma dei quadrati degli scostamenti dei termini in confronto d'ogni altra costante, introduce due nuovi concetti nella teoria statistica: il concetto dello *scostamento medio* e quello della *precisione* delle serie.

---

(1) Ricordiamo che il seno di un angolo non è altro se non il rapporto tra la perpendicolare opposta all'angolo e l'ipotenusa o raggio, fatto eguale all'unità. Nella figura qui accanto, se l'angolo *a c d* è di 30 gradi, la perpendicolare *a b* è la metà esatta della ipotenusa o raggio *a c*; ossia *sen* 30° = 0,5; se l'angolo è di 31°, la perpendicolare *a'b'* riesce un po' più di 51 ½ centesimi dell'ipotenusa o raggio *a'c* (e più precisamente: *sen* 31° = 0,51504...).

Se chiamiamo $x_{,}, x_{,,}, x_{,,,} \dots x_n$ gli scostamenti semplici dei termini di una serie della loro media aritmetica, si avrà per le cose dette:

$$x_{,}^2 + x_{,,}^2 + x_{,,,}^2 + \dots + x_n^2 = minimum.$$

Sia μ^2 la media aritmetica di cotesti quadrati; sarà:

$$\mu^2 = \frac{x_{,}^2 + x_{,,}^2 + x_{,,,}^2 + \dots + x_n^2}{n}$$

perciò

$$\mu = \sqrt{\frac{x_{,}^2 + x_{,,}^2 + x_{,,,}^2 + \dots + x_n^2}{n}} = \sqrt{\frac{[x^2]}{n}}$$

dove il simbolo della parentesi quadra esprime in forma abbreviata la somma dei quadrati dei diversi x. Ora il valore di μ è ciò che dicesi « scostamento medio » (1).

La « precisione » di una serie può definirsi il rapporto fra lo scostamento medio e la media aritmetica, cioè:

$$\frac{\sqrt{\frac{[x^2]}{n}}}{m} = \frac{\sqrt{[x^2]}}{m\sqrt{n}}$$

Quanto più piccolo è lo scostamento medio rispetto alla media aritmetica, tanto maggiore la precisione della serie, ossia tanto più ristretto è il *campo di variabilità* dei termini che compongono la serie.

Di due serie d'egual numero di termini e della stessa media aritmetica, si dirà più precisa quella che ha lo scostamento medio minore.

Di due serie di differente numero di termini e di differente media aritmetica, si dirà più precisa quella che presenta un rapporto più piccolo tra lo scostamento medio e la media aritmetica.

§ 9. Gli statistici distinguono ancora le medie in *oggettive* e *soggettive*, *tipiche* e *non tipiche*, *semplici* e *ponderate;* distinguono pure media da *valor mediano* e da *norma* o *valore di densità*.

Media « oggettiva » (detta anche *reale*) è quella desunta da osservazioni ripetute intorno ad una medesima grandezza, la quale esiste in realtà con dimensioni definite, ma che i nostri sensi o strumenti imperfetti non riescono a misurare con precisione assoluta. Ne abbiamo dato testè un esempio. Media soggettiva (o anche *fittizia*, *ipotetica*, o *media di semplice perequazione*) è quella desunta da osservazioni esatte o ritenute esatte riferibili a grandezze diseguali e reali della stessa specie. Così, se in un anno sono registrate 1010 nascite, in un secondo 950, in

(1) Lo si chiama anche *scostamento quadratico medio* per distinguerlo dallo *scostamento semplice medio*, il quale si ottiene facendo la media aritmetica degli scostamenti semplici, considerati questi come se fossero tutti di segno positivo

un terzo 1028, la media aritmetica, che è 996, si dirà fittizia o astratta (o impropriamente soggettiva), come quella che indica in quale misura si sarebbe verificato il fenomeno della natalità, se si fosse distribuito uniformemente nel tempo.

Media « tipica » è quel valore intorno al quale si addensano maggiormente i dati delle varie osservazioni, con distribuzione piuttosto simmetrica e con frequenza tanto minore quanto più ci allontaniamo dalla media stessa, in un senso o nel senso opposto. Così distribuendo 10.000 italiani secondo l'indice cefalico, si avrebbe intorno alla media 82,7 una ripartizione abbastanza simmetrica, una frequenza di casi tanto più decrescente quanto più dalla media ci discostiamo per andare verso gli indici cefalici più bassi o verso i più elevati. Questa media si dice perciò tipica, come quella che allude ad un concorso di fattori etnici, sociali, ecc., che ha avuto per risultato un *maximum* di frequenza dei casi in corrispondenza di essa media; mentre i casi corrispondenti ad altri valori possono considerarsi come deviazioni dal tipo, quasi errori di costruzione e di riproduzione di un modello prestabilito da natura.

Indice cefalico	Individui osservati
— 70 per cento	12
70-72 »	103
73-75 »	463
76-78 »	1.261
79-81 »	2.020
82-84 »	2.702
85-87 »	1.898
88-90 »	1.086
91-93 »	347
94-96 »	96
Oltre 96 »	12
	10.000

Medie « quasi tipiche » sono poi quelle dedotte da seriazioni, che soddisfano imperfettamente alle condizioni predette, di solito a causa dell'asimmetria nella distribuzione dei casi, per cui la media calcolata non può coincidere col punto di massimo addensamento. Sono infine « non-tipiche » le medie che addirittura corrispondono nella seriazione a punti di minimo addensamento. Così l'età media dei morti in Italia è di 27 anni circa, cioè corrisponde ad un'epoca della vita, in cui la mortalità assoluta o relativa, lungi dall'essere massima, è quasi al suo minimo. Le massime mortalità assolute si hanno nei primi anni dalla nascita e intorno al 72° anno. In generale sono non-tipiche le medie calcolate su seriazioni che hanno due massimi agli estremi.

I punti di maggior addensamento delle seriazioni, che il Bowley chiama « *mode* », li chiameremo per intenderci « norme ». La seriazione sarà *uninormale*, se presenta un solo punto di massimo addensamento; *plurinormale*, se ne presenta due o più.

« Valor mediano » dicesi quel punto della scala di misura, in corrispondenza del quale si bipartisce esattamente la seriazione. Se disponiamo in ordine decrescente di stature 101 coscritti, la statura del 51[mo] costituisce il valor mediano del gruppo.

« Media semplice » è quella che si ricava da osservazioni aventi egual peso od importanza; « media ponderata » quella nel calcolo della quale ogni elemento entra per tante volte, quante ne richiede il peso o grado suo d'importanza, ossia tante volte quanti i casi in cui fu osservato. Così se sopra un mercato si son venduti 100 quintali di grano al prezzo unitario di lire 20; poi altri 150 al prezzo di lire 20,70; poi ancora altri 300 al prezzo di lire 22, non sarebbe logico concludere che il prezzo medio è $= \frac{20 + 20,70 + 22}{3} = 20,90$, perchè la partita di grano che fu venduta a lire 22 era tre volte più importante di quella che fu venduta a 20 e due volte più di quella che ottenne il prezzo di 20,70. Bisogna invece calcolare il prezzo complessivo di vendita di tutte tre le partite e dividere per il numero complessivo di quintali, ossia moltiplicare le singole osservazioni (prezzi unitari) per i rispettivi pesi o coefficienti d'importanza (quintali venduti), fare la somma dei prodotti e dividerla per la somma dei pesi o coefficienti:

$$\frac{20 \times 100 + 20,70 \times 150 + 22 \times 300}{100 + 150 + 300} = \text{L. } 21,28.$$

In termini generali, essendo $a_1, a_2, a_3, \ldots a_n$ gli elementi osservati e $p_1, p_2, p_3 \ldots p_n$ i rispettivi pesi o coefficienti o casi osservati, si avrà come espressione della media ponderata:

$$M = \frac{a_1 p_1 + a_2 p_2 + a_3 p_3 + \ldots + a_n p_n}{p_1 + p_2 + p_3 + \ldots + p_n}.$$

Nella seriazione degli indici cefalici data a pagina precedente, gli elementi osservati sono gli indici cefalici stessi; i pesi o coefficienti d'importanza sono gl'individui. Però siccome gli elementi sono dati in classi (70-72, 73-75, 76-78, ecc.), il calcolo della media ponderata esige qualche artificio. Così si può prendere, nella classe 70-72, il valore 71 come centrale e supporre che i 103 individui avessero tutti un indice cefalico precisamente di 71; e del pari nelle classi 73-75, 76-78, 79-81, ecc., prendere come centrali i valori 74, 77, 80, ecc., e supporre ancora che gli individui appartenenti a dette classi avessero tutti l'indice di 74 o di 77 o di 80, ecc., rispettivamente. Il calcolo della media vien così molto agevolato; però nel caso concreto non sarebbe molto rigoroso. Infatti siccome in questa prima parte della seriazione i numeri vanno crescendo da classe a classe, è probabile che essi si troverebbero crescenti anche da gruppo a gruppo di una stessa classe. Ad es.: è probabile che gli individui con un indice di 71 siano in maggior numero di quelli di 70; e quelli con un indice di 72,

in maggior numero di quelli del 71. Per il rigore del metodo si richiederebbe dunque di precisare in ogni classe il valore dell'indice, che bipartisce esattamente il numero dei soggetti osservati in essa classe; al quale scopo giovano i procedimenti d'interpolazione. Tuttavia nelle seriazioni che presentano una certa simmetria, l'adottare il punto di mezzo tra i limiti indicati delle classi non nuoce gran fatto alla media generale ponderata, perchè gli errori in difetto, che si commettono nella prima parte della seriazione, si compensano con quelli in eccesso della seconda parte.

B) *Proporzioni e rapporti.*

Sommario: § 1. Rapporti che si semplificano e rapporti che si risolvono colla determinazione del quoziente. — § 2. A) Rapporti di ***composizione* o *di parte al tutto*** - B) Rapporti di *derivazione semplice* - C) di *derivazione complessa* - D) di *effetto a causa* - E) di *durata* - F) di *ripetizione*. — § 3. Argomenti rinviati ad altra sede.

§ **1.** Del concetto di *rapporto* si è già parlato in sede di nozioni generali introduttive. Qui ci occuperemo delle proporzioni e dei rapporti, come procedimenti aritmetici semplificativi e rappresentativi delle quantità rilevate.

Abbiamo anche detto che i rapporti statistici non si possono istituire indifferentemente fra termini presi ad arbitrio; ma occorre, a fondamento del processo aritmetico, una relazione logica come di parte al tutto, di contenuto a contenente, di effetto a causa, di situazione statica a dinamica, di manifestazione accidentale a manifestazione normale, e così via. Solo allora dal confronto delle due quantità potremo derivare un concetto nuovo, che sarà appunto, secondo i casi, quello della relativa importanza o frequenza, della probabilità, della durata, dell'anormalità, ecc., del fatto in esame.

Non è necessario che i termini di un rapporto statistico siano entrambi quantità statistiche; basta che lo sia uno dei due. Nel calcolo della *densità della popolazione*, la superficie del territorio, cui la popolazione vien ragguagliata, è un dato puramente geografico, una quantità fissa, non dell'ordine dei fenomeni variabili di cui si occupa la statistica. La popolazione, sì, appartiene a quest'ordine. Così, se confrontiamo i casi osservati di parti gemelli composti di due maschi o di due femmine o di un maschio e di una femmina coi casi che si dovrebbero avere in una combinazione a sorte, il rapporto si stabilisce fra un dato statistico e uno desunto dal calcolo di probabilità; però il rapporto stesso non cessa di essere statistico, perchè la presenza di un elemento statistico gli conferisce quel carattere di variabilità senza regola assegnabile a tutto rigore, che esige una speciale metodologia.

I numeri assoluti, in cui sono espressi i termini originarii in paragone tra loro, presentano quasi sempre grandezze comparative male

afferrabili dal nostro spirito; donde l'opportunità della « semplificazione » del rapporto in certi casi e della sua « risoluzione » (mediante la determinazione del quoziente) in certi altri.

Operiamo la *semplificazione*, quando entrambi i concetti, cui si riferiscono le quantità confrontate, debbono essere tenuti presenti e distinti nell'enunciato finale. Uno dei termini si fa eguale all'unità o a un numero rotondo, di solito una potenza intera del 10, come 100, 1000..... 1 milione, e l'altro si riduce in proporzione secondo la nota regola del tre. Ad es., sapendosi che nel 1899 si ebbero in Italia 45.910 nati-morti su 1.134.468 nascite in generale (nati-vivi e nati-morti insieme) si stabilisce la proporzione:

$$1.134.468 : 100 = 45.910 : x$$

$$x = 4{,}05 \text{ (più precisamente } 4{,}04683...).$$

Il rapporto dei nati-morti alle nascite, espresso in forma semplice e ritenibile dalla memoria, è dunque $\frac{4{,}05}{100}$, che può scriversi: 4,05 %. La semplicità si è ottenuta però con leggiero scapito della precisione, essendosi dovuto per quella rinunziare agli ultimi decimali del numeratore. Come ben si vede, i due concetti di *nati-morti* e di *nascite in generale* sono ancor tenuti presenti e distinti nel nuovo enunciato; il primo vi figura colla cifra 4,05; il secondo colla cifra 100 o col simbolo %.

Invece il rapporto non si semplifica, ma si *risolve* mediante la ricerca del quoziente, allorchè questo assume un significato suo proprio, diverso dal significato del dividendo e da quello del divisore. Così, se divido l'ammontare dei depositi a risparmio, esistenti presso una Banca, per la semisomma dei versamenti e rimborsi dell'anno, ottengo una cifra che esprime la *durata* media di ogni deposito in anni o frazioni di anno. Qui dunque i termini originarii del rapporto non sono più tenuti presenti e distinti, come poc'anzi, nell'enunciato finale; ma si son risolti in un dato nuovo, di significato suo proprio; da idee di « valore » siamo passati a quella di « tempo ».

In generale, la semplificazione si desidera nei rapporti di composizione, di derivazione causale e simili; la determinazione del quoziente o risoluzione del rapporto è necessaria invece tutte le volte che si può e si vuole dai termini in confronto ricavare il concetto della *durata* di un fenomeno o quello della sua *ripetizione* sotto date forme, statisticamente osservabili in successivi tempi.

Qualora si abbia per una stessa categoria di fatti una serie di rapporti da semplificare, conviene che il termine di riferimento ridotto a 100 o a 1000, ecc. sia sempre il medesimo per tutti i gradi della serie; questo termine vien considerato come *fisso*, l'altro come *variabile*. Ma quale dei due si considererà variabile? Dal punto di vista aritmetico la cosa è indifferente; dal punto di vista logico-statistico, no. Se, ad es., esamino per varii anni la frequenza di certi delitti in una popolazione,

logicamente assumerò come termine variabile il numero dei delitti, perchè è quello appunto di cui voglio conoscere, in numeri semplici e facili alla memoria, le oscillazioni annuali; e ridurrò a cifra rotonda e fissa il numero degli abitanti. In altre parole, cercherò quanti delitti di quella data specie si ebbero ogni 1000 abitanti, anno per anno; non quanti abitanti si noverano per ogni delitto. Aggiungasi come motivo d'opportunità questo, che se si prendesse la popolazione come termine variabile, il crescere del numero degli abitanti per ogni delitto potrebbe, per facile equivoco, interpretarsi come aumento di delinquenza, laddove in realtà significherebbe il contrario. Il nostro intelletto corre più spontaneamente all'idea di una proporzione diretta che a quella di una inversa.

In generale, dunque, dovendosi semplificare un rapporto, conviene assumere come elemento variabile quello appunto di cui c'interessano le variazioni o che è già per propria natura il più variabile; quindi si tratterà come tale la parte rispetto al tutto, il contenuto rispetto al contenente (es.: la popolazione rispetto al territorio), il caso particolare o anormale rispetto al normale o medio, l'effetto rispetto alla causa, ecc.

Naturalmente questi criterii non hanno applicazione in quella classe di rapporti, che non si semplificano, ma si risolvono colla determinazione del quoziente.

§ **2.** Passiamo in rapida rassegna le principali specie di rapporti:

A) *Rapporti di composizione o di parte al tutto.* — Sono i più elementari e comuni. Esempi: rapporto tra il numero dei giorni piovosi e quello totale dei giorni dell'anno; tra il numero degli emigranti in età inferiore a 15 anni e quello degli emigranti in complesso, senza distinzione d'età; tra la quantità venduta di sale raffinato e quella dei sali commestibili d'ogni specie, ecc.

La sola avvertenza, che vuolsi fare qui, è che allorquando nella totalità delle osservazioni entrano elementi non specificati, conviene piuttosto ragguagliare la *parte nota* alla *somma delle parti note*, deducendo cioè dal totale la categoria degli elementi, che non si son potuti classificare. Poniamo: nel 1897 si noverarono in Italia 478 suicidii per arma da fuoco su un totale di 1895 suicidii, di cui però 180 compiuti con mezzi non specificati. È chiaro che il rapporto non deve stabilirsi fra 478 e 1895, ma fra 478 e (1895-180), ossia $\frac{478}{1715}=\frac{27,9}{100}$. Peraltro se vi sono indizi, in base ai quali ripartire gli elementi non specificati fra le categorie stabilite, convien farlo anche in via di semplice approssimazione, o almeno convien fare avvisato il lettore che il rapporto calcolato tra la parte nota e la somma delle parti note deve ritenersi, secondo i casi, inferiore o superiore al vero. Così se la statistica ci informasse che i 180 suicidii, di cui non fu precisato il mezzo, si riferiscono in gran parte a donne, oppure anche a uomini, ma di età avan-

zata, tra cui più raro è l'impiego dell'arma da fuoco pel triste proposito gioverebbe avvertire che il rapporto $\frac{27,9}{100}$ è superiore al vero, anzichè inferiore.

Un rapporto di composizione è anche quello che si istituisce fra una parte e l'altra, fra una categoria e l'altra di un medesimo fatto statistico. Ma se eccettuiamo il caso delle nascite maschili, che si sogliono ragguagliare alle femminili (x maschi per 100 femmine, anzichè x maschi per 100 nati di qualunque sesso), questa forma di rapporto è raramente usata.

B) *Rapporti di derivazione semplice.* — Occorre spesso di dover determinare la frequenza di un fenomeno rispetto ad un altro che ne è, per dir così, la condizione generale di esistenza o che gli fornisce gli elementi perchè si possa produrre. Nell'ordine sociale è la popolazione quella che appresta materia alla realizzazione d'innumerevoli fenomeni statistici; sicchè ad essa usiamo ragguagliare matrimoni, nascite, morti, migrazioni, criminalità, valore del commercio estero, ammontare delle imposte, ecc. Tutti questi rapporti esprimono la *frequenza generica* dei fenomeni, della cui esistenza è condizione generale la popolazione; la loro utilità dipende da ciò che i numeri assoluti, in cui è espresso un fenomeno, per se stessi non dicono niente di preciso sull'importanza del fenomeno medesimo. Sapere che gli Italiani pagano 1500 milioni d'imposta e i Francesi 4000, non significa ancora che *un* cittadino italiano paghi in media meno di un cittadino francese; per decidere intorno a ciò bisogna conoscere o la popolazione totale o il totale reddito dei due paesi. La popolazione riconosce alla sua volta nel territorio una condizione generale di esistenza. Il rapporto tra quella e la superficie indica la frequenza media o *densità* per l'unità areale adottata. La superficie diventa poi il termine naturale di confronto dei fenomeni, di cui si vuol mettere in evidenza l'estensione geografica; così si dirà: tanti ettari coltivati a grano ogni km²; ovvero tanti chilometri di ferrovia ogni 100 km², ecc.

Noi possiamo inoltre scegliere in seno alla popolazione o nell'àmbito del territorio elementi che siano condizione più diretta dei fatti in esame; e allora sorge il concetto di *frequenza specifica.* Se invece di riferire le nascite legittime a tutta intera la popolazione, le riferisco alla sola classe delle donne maritate o, meglio ancora, alla classe delle maritate in età presumibilmente feconda, tra i 15 e i 50 anni, avrò la frequenza specifica delle nascite legittime. La superficie coltivata a grano si potrà ragguagliare, anzichè a tutto il territorio, alla sola area suscettiva di colture agrarie, esclusi quindi i terreni assolutamente improduttivi (alta montagna), i bacini e corsi d'acqua, aree fabbricate, ecc.

Qui giovano alcuni avvedimenti per non cadere nell'errore di assumere come completo il termine di riferimento, mentre può non esserlo. Le tavole di mortalità ci forniscono l'illustrazione migliore del caso.

La frequenza specifica della mortalità determinavasi un tempo dividendo senz'altro il numero dei *morti* d'ogni età per quello dei *viventi della stessa classe d'età*. Ma il denominatore era evidentemente incompleto, perocchè a formare i casi possibili di mortalità hanno concorso, non i soli individui che a un certo istante si trovano presenti e vivi in una data classe annuale d'età, ma tutti coloro che nel corso di dodici mesi hanno fatto il loro ingresso in quella classe; il che è diverso. Per chiarire la cosa con un facile paragone, diremo: come le spese di un'annata non si debbono ragguagliare a ciò che residua in cassa ad un dato momento, ma a ciò che è *entrato* in cassa nel corso dei dodici mesi, così i deceduti nell'anno debbono essere ragguagliati agli entrati vivi nella classe d'età in questione, non a coloro che a un momento dato residuano viventi. Ora fra gli entrati vivi contano per qualche cosa i morti medesimi. Infatti per figurare come deceduti in una qualsiasi classe d'età, bisogna ben esserci entrati vivi e rimasti anzi presenti in media per circa sei mesi, perocchè la scomparsa di un gruppo d'individui non avviene per tutti subito dopo incominciato il nuovo anno d'età, nè per tutti ad eguale distanza da questo punto. Ecco perchè conviene aggiungere ai viventi di una data classe d'età all'incirca la metà dei morti, ritenendosi che questi siano stati presenti e vivi in detta classe in media per metà dell'anno. La necessità di questa correzione, per la quale al concetto di « viventi » si viene a sostituire nel denominatore del rapporto il concetto di « esposti a morire », salterebbe meglio all'occhio, se considerassimo i morti da 0 a 1 anno. Questi, non c'è dubbio, vanno ragguagliati ai *nati*, non già ai bambini esistenti da 0 a 1 anno, che sono in qualsiasi istante in numero minore dei nati nel corso dell'anno, a causa delle perdite inflitte dalla morte.

Per parità di ragionamento diremo che i morti da 1 a 2 anni, da 20 a 25, da 30 a 60, da zero a ω, vanno ragguagliati, non già ai viventi da 1 a 2, da 20 a 25, da 30 a 60, da zero a ω, ma alla somma degli entrati vivi nei singoli anni delle classi di età anzidette, la qual somma è a un dipresso eguale a quella dei viventi accresciuti della metà dei morti. Dunque, anche quando si cerca la frequenza generica della mortalità in una popolazione, si dovrà stabilire il rapporto:

$$\frac{m}{P + \frac{m}{2}}$$

e non quello comunemente usato $\frac{m}{P}$ (1).

(1) Le stesse considerazioni vogliono essere tenute presenti in casi analoghi, per esempio, nella determinazione del saggio della *nuzialità specifica*. Il numero degli sposi si ragguaglierà a quello dei celibi da 18 anni in su aumentati della metà degli sposi.

Altri avvedimenti mirano ad assicurare sotto diversi aspetti l'omogeneità dei termini del rapporto. Nel lasso di tempo che suol esserci tra il fatto statistico derivato e quello che per dir così lo alimenta, possono verificarsi mutazioni non trascurabili. Talvolta si trascurano e si presumono compensate le variazioni in aumento e quelle in diminuzione. Così nel calcolo di sopravvivenza dei bambini dalla nascita ai 5 anni trascuriamo le perturbazioni dovute ai movimenti migratorii. Ma se la mutazione non può avvenire che in un senso soltanto, è consigliabile di calcolarla o almeno tenerla presente. Caratteristico ci si offre il caso del rapporto tra i morti legittimi da 0 a 1 anno d'età e i nati legittimi, poichè tra i primi possono figurare dei bambini, che nacquero *illegittimi*, furono *legittimati* dopo qualche mese per susseguente matrimonio dei genitori e morirono *legittimi* nell'anno stesso. L'errore si farebbe più grande nel confronto dei bambini legittimi morti da 1 a 2 anni, da 2 a 3, ecc., coi superstiti legittimi a 1 anno, a 2 anni, ecc., calcolati col metodo di Hermann, di cui parleremo, e senza le correzioni fornite da una statistica delle legittimazioni.

C) *Rapporti di derivazione complessa.* — La frequenza di un fatto statistico rispetto ad un altro, da cui trae gli elementi, può calcolarsi in due modi: o ragguagliando il *movimento* del primo allo *stato* numerico medio del secondo (nel qual caso abbiamo un rapporto di derivazione semplice); ovvero ragguagliando il *movimento* del primo al *movimento corrispondente* del secondo. In questo secondo caso il rapporto può chiamarsi di « derivazione complessa », come quello che involge insieme uno speciale concetto di tempo.

Se diciamo, ad esempio: in Italia si rimaritano ogni anno 14.500 vedove su un totale di 1.400.000 vedove esistenti e disponibili, ossia 1,04 %, questa è la percentuale di vedove esistenti (d'antica o di recente vedovanza), che ripassano allo stato coniugale *nel corso di un anno*. Non vi può essere dubbio che tale sia il significato del rapporto. Ma se diciamo: in Italia si rimaritano ogni anno 14.500 vedove su 96.000 che « diventano » vedove ogni anno, questo rapporto, che semplificato è pari a 15,1 %, indica *quante delle coniugate divenute vedove nell'anno passeranno a nuove nozze nel corso della restante loro vita*.

Soffermiamoci sulla interpretazione del secondo rapporto, perchè quella del primo è per se stessa evidente.

Supponiamo per comodità di ragionamento che la popolazione sia *stazionaria* in tutta la sua struttura demografica. In tale ipotesi è chiaro che le donne, la cui vedovanza risale, per esempio, all'anno scorso e che si rimaritano in quest'anno 1904, debbono essere eguali in numero a quelle che, vedovate nel 1904, riprenderanno marito l'anno venturo; e similmente che le donne, la cui vedovanza risale a due anni fa e che celebrano seconde nozze in quest'anno 1904, debbono egua-

gliare per numero quelle che, vedovate nel 1904, si rimariteranno fra due anni; e così via. Si arriva insomma a concludere che le vedove rimaritatesi nel 1904, *a qualunque epoca risalga la loro vedovanza*, pareggiano in numero quelle che, diventate vedove nel 1904, convoleranno a nuove nozze negli anni avvenire. Perciò a chi domandasse: « quante delle vedove, divenute tali nel 1904, si rimariteranno in qualunque degli anni avvenire », la risposta dovrebbe essere questa: « tante, quante sono le vedove, d'antica o di recente vedovanza, che si sono nuovamente coniugate nel 1904 » (1).

Se la popolazione non è stazionaria, ma progressiva, allora è chiaro che le vedove, rimaritantisi oggi e la cui vedovanza risale a varie date, provengono da generazioni meno numerose di quelle delle donne vedovate quest'anno e che si consoleranno con nuovo imeneo negli anni venturi. In tal caso, per ottenere la miglior corrispondenza di tempo fra i due termini del rapporto, bisogna ragguagliare le seconde nozze di vedove, per un dato periodo, alle vedovanze verificatesi in un periodo anteriore a quello di tanto, quant'è la durata media delle vedovanze temporanee.

D) *Rapporti di effetto a causa.* — Sono una varietà dei rapporti di derivazione semplice o complessa. Esempio tipico: le nascite legittime rispetto ai matrimoni. Qui pure il confronto può farsi tra il movimento di un fenomeno e lo stato medio dell'altro, oppure tra movimento e movimento.

In Italia su 4.371.393 coniugate in età tra 15 e 50 anni, censite il 10 febbraio 1901, si ebbero, nel 1902, 1.030.543 nati vivi legittimi, ossia 236 nati per 1000 coniugate.

Gli stessi nati, in numero di 1.030.543, riferiti invece alle 237.513 donne che si sposarono nel 1902, dànno per quoziente 4,39, il quale

---

(1) La cosa si può rendere più chiara con un semplice schema. Supposta come dicemmo, la popolazione stazionaria, si deve avere:

	Donne vedovate nel	che passano a nuove nozze nel	sono eguali per numero	alle donne vedovate nel	che si rimariteranno nel
	1903	1904	=	1904	1905
	1902	1904	=	1904	1906
	1901	1904	=	1904	1907
	. . .	. .	. . .	. . .	. . .
	. . .	. . .	. . .	. . .	. . .
	. . .	. . .	. . .	. . .	. . .
	1885	1904	=	1904	1923
	ecc.	ecc.		ecc.	ecc.
Totali . . .	1903 in addietro	1904	=	1904	dal 1905 in avanti

quoziente significa che la fecondità media d'ogni matrimonio, distribuita su tutto il periodo della convivenza feconda, è di 4,39 figli.

Come si spiega che le nascite di un anno, le quali provengono da matrimoni contratti non solo nell'anno stesso, ma uno, due, tre o più anni addietro, possono considerarsi come frutto dei matrimoni d'un medesimo anno per tutta la durata loro?

La risposta è conforme a quella data poc'anzi nell'esempio delle vedovanze e delle seconde nozze. Supposta la popolazione stazionaria in tutta la sua struttura, supposizione che per il momento semplifica il problema, è ovvio che i nati quest'anno da matrimoni contratti un anno fa, saranno ugualmente numerosi dei nascituri fra un anno dai matrimoni d'oggi; e così pure, che il numero dei nati quest'anno da matrimoni contratti due, tre o più anni addietro deve pareggiare quello dei nascituri fra due, tre o più anni dai matrimoni d'oggi. Dunque il totale dei nati attualmente da matrimoni che risalgano a qualunque epoca eguaglia il totale dei nascituri in tutti gli anni avvenire dai matrimoni di quest'anno (1).

Se la nuzialità non fosse stazionaria, ma da lungo tempo in aumento, e la natalità pure, allora i nascituri dai matrimoni attuali dovrebbero prevedersi più numerosi dei nati attualmente dai matrimoni che risalgono a diverse date nel passato; e per ottenere una meno imperfetta corrispondenza di tempo fra i due termini del rapporto, converrebbe dividere le nascite di un certo periodo per i matrimoni di un periodo anteriore di tanto, quant'è in media il tempo che le coppie coniugali impiegano per mettere al mondo la metà delle loro figliuolanze.

E) *Rapporti di durata.* — Questi rapporti appartengono alla categoria di quelli che non si semplificano, ma si risolvono con la determinazione del quoziente. S'istituiscono fra la *consistenza* nume-

---

(1) Anche qui un semplice schema può rendere più chiara la soluzione del problema. Supposta la popolazione stazionaria si deve avere:

I nati nell'anno	da matrimoni contratti nel	debbono essere eguali per numero	ai nascituri nell'anno	da matrimoni contratti nel
1904	1903	=	1905	1904
1904	1902	=	1906	1904
1904	1901	=	1907	1904
1904	1900	=	1908	1904
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...
1904	in qualunque epoca a risalire dal 1903	=	in qualunque epoca dal 1905 in poi	1904

rica di un fenomeno statistico e il suo *movimento di rinnovazione e di estinzione*. La durata è espressa dal quoziente in anni, mesi, ecc., secondo che il movimento del fenomeno è considerato per lo spazio di un anno, di un mese, ecc.

Facciamo il caso di una Banca il cui portafoglio consti in media di 12 milioni di lire in effetti commerciali. Di questi se ne estinguano per 3 milioni ogni mese. Supponiamo pure una condizione di cose stazionaria e cioè che anche le nuove cambiali ammesse allo sconto, che prendono man mano il posto delle estinte, ammontino a 3 milioni al mese. Dico che la *scadenza media* degli effetti in portafoglio è di tanti mesi, quante volte la cifra di 3 milioni sta in quella di 12 milioni, ossia 4 mesi. È ovvio infatti che per effetto delle nuove cambiali surroganti le estinte il portafoglio si rinnova ogni mese per una quarta parte del suo ammontare; quindi si troverà rinnovato per intero in 4 mesi; che è come dire che ogni titolo, in esso compreso, non sarà surrogato da uno nuovo se non in capo a 4 mesi in media. Dunque la durata media di presenza d'ogni effetto nel portafoglio della Banca sarà di 4 mesi.

Se però la situazione della Banca non fosse stazionaria, ma, per esempio, in progresso continuo e quindi ogni nuova partita di effetti scontati superasse di regola la partita mensile estinta, è ovvio che la surrogazione o rinnovazione completa del portafoglio verrebbe affrettata e la scadenza media di ogni effetto risulterebbe minore di quattro mesi.

Come dunque dovrebbe in tale ipotesi impostarsi il rapporto? Evidentemente le cambiali, che i debitori della Banca estinguono oggi, vanno riferite non alla consistenza attuale del portafoglio, che formossi in grazia di sconti recenti e più numerosi di quelli in corso di estinzione, ma alla consistenza qual'era alcuni mesi addietro, allorchè le cambiali oggi estinte erano giunte a metà del periodo di loro presenza in portafoglio e questo poteva ritenersi composto in eque parti di effetti scontati anteriormente e in minor numero di quelli che sono ora in corso di estinzione e di effetti scontati successivamente e in maggior numero.

In ultima analisi torna lo stesso che si divida la consistenza del portafoglio di alcuni mesi addietro (quanti equivalgono alla metà del periodo di giacenza) per le cambiali estinte nell'ultimo mese, e che si divida la consistenza attuale per la semisomma delle cambiali estinte e delle nuove ammesse allo sconto.

Per parità di ragionamento si concluderebbe che la durata media di una *generazione* si ottiene dividendo la popolazione per la semisomma delle nascite e delle morti annuali; la durata media della *convivenza coniugale*, dividendo il numero delle coppie coniugali esistenti per la semisomma delle coppie annuali nuove di sposi e di quelle disciolte per morte di uno dei coniugi; la *giacenza media* dei depositi

presso una Cassa di risparmio, divivendo l'ammontare dei depositi esistenti per la semisomma dei versamenti e dei rimborsi, e così via.

Di siffatti rapporti le applicazioni sono frequenti; il loro grande vantaggio sta nel risparmiare una rilevazione diretta, che riuscirebbe lunga e costosa. Un istituto di credito popolare, le cui operazioni passive consistano soprattutto di depositi a risparmio e in conto corrente, e le attive di sconti e anticipazioni, deve regolarsi in modo che la scadenza delle seconde rimanga d'un certo tratto al di sotto della durata media dei depositi. Elevare l'interesse sui depositi equivarrà in generale ad allungare la durata dei depositi; elevare il saggio dello sconto equivarrà ad abbreviare la scadenza media delle cambiali, almeno nel senso di diminuire la proporzione in portafoglio delle cambiali a scadenza lunga. Così, una Banca di emissione che lavori in concorrenza con altre, ha ragion di temere che i proprii biglietti siano male accetti al pubblico, se il periodo medio, in capo al quale rientrano in cassa, si accorcia tendendo al limite di scadenza delle operazioni attive a più breve termine, e se per le Banche rivali, a parità d'altre circostanze, succede invece il contrario.

Poichè siamo in argomento, ci sembra opportuno ricordare il metodo col quale il Des Essars calcola la velocità di circolazione della moneta nel caso particolare dei conti correnti. Il metodo non è sostanzialmente diverso da quello su indicato; solo che mette in risalto la *velocità*, e non la *durata*.

« È noto, egli dice (1), che i commercianti e molti privati hanno il costume di depositare i loro fondi disponibili presso le Banche e di metterli in moto per mezzo di *buoni di cassa* e *chèques*... I versamenti fatti dai titolari dei conti correnti formano il « credito », i pagamenti delle Banche eseguiti per ordine dei titolari sui conti correnti stessi formano il « debito » dei correntisti. Il « saldo » consiste nella differenza fra debiti e crediti.

« I conti correnti, così continua il Des Essars, sono esattamente paragonabili a un serbatoio che riceva un liquido (credito) e lo lasci defluire per un orificio (Banca). I saldi sono il livello che il liquido ha nel serbatoio e il debito è la parte scolata via.

« Per calcolare la velocità del corpo di liquido che passa per l'orificio, immaginiamo che un debitore *d* paghi un creditore *c;* l'operazione consiste nel trasportare una certa quantità di numerario o di qualsiasi altro mezzo di pagamento da *d* a *c*.

d ——————— *b* ——————— *c*

« Chiamo 1 la distanza *dc*. Se interpongo una Banca *b* a eguale distanza da *d* a *c*, il debitore farà percorrere al numerario la distanza

(1) V. *Giornale degli Economisti*, maggio 1895.

db, cioè $^1/_2$ e la Banca farà percorrere a questo stesso numerario la distanza *bc*, cioè $^1/_2$.

« I clienti di una Banca versano annualmente presso di essa la somma m, totale dei crediti; la Banca paga, pure annualmente la somma m', in generale diversa da m, onde lo spostamento totale della massa monetaria per mezzo dei conti correnti ha per espressione:

$$\frac{m + m'}{2}.$$

« Questo spostamento totale (è sempre il Des Essars che parla) non è altro che la somma degli spostamenti giornalieri dei saldi detenuti dalle Banche, o, più semplicemente, lo spazio percorso annualmente dal saldo medio. Quindi, indicando con s il saldo medio o consistenza media e con v la velocità annuale, si ha la relazione:

$$v \, . \, s = \frac{m + m'}{2},$$

il che in linguaggio ordinario si traduce così: *la semisomma annuale dell'avere e del dare dei conti correnti è eguale alla quantità del movimento del saldo medio dell'anno* ».

Lo scrittore francese calcolando in base ai dati delle grandi Banche europee la velocità dei conti correnti, secondo la formola derivata dalla precedente:

$$v = \frac{m + m'}{2s}$$

trova che la velocità tocca sempre un massimo nelle epoche di crisi e un minimo nelle epoche di liquidazione. Egli trova pure che nei paesi a finanze avariate la velocità di circolazione soffre di una riduzione enorme. La velocità dei conti correnti sarebbe così un prezioso barometro della situazione del mercato e un indice dello stato delle finanze.

Ora, senza entrare nella critica di queste illazioni (1), vogliamo solo dimostrare che quella formola non è che una variante della nostra. Noi avremmo calcolato la *durata di giacenza* dei depositi in

---

(1) Solo a riguardo della conclusione affermata rispetto ai paesi dalle finanze avariate, esponiamo il dubbio che vi sia una relazione così diretta fra la situazione finanziaria e la velocità dei conti correnti. A noi pare piuttosto che i paesi dalle finanze avariate, cui si riferisce il Des Essars, siano di quelli industrialmente poco evoluti, le cui produzioni hanno carattere prevalentemente agricolo e perciò sono più vincolate alle stagioni (raccolti). Questo vincolo alle stagioni fa sì che incassi e pagamenti si verifichino in modo più discontinuo che non nei paesi che hanno un corredo più completo di industrie di vario genere; la qual cosa non può non sortire l'effetto di un prolungamento nelle giacenze di depositi. Nè si dica che questo effetto si limiterà ai depositi fruttiferi; poco o molto debbono risentirsene anche gli

conto corrente dividendo la consistenza loro (saldo medio) per la semisomma dei versamenti e dei ritiri, degli accreditamenti e degli addebitamenti. Chiamando t il tempo di giacenza, s la consistenza o saldo medio, m gli accreditamenti (o versamenti), m' gli addebitamenti (o ritiri o pagamenti fatti dalla Banca su ordine dei correntisti), avremmo avuto:

$$t = \frac{s}{\frac{m + m'}{2}}$$

donde con facili trasformazioni:

$$\frac{1}{t} = \frac{m + m'}{2s}.$$

Ora si vede subito che il simbolo v usato dal Des Essars corrisponde perfettamente al nostro $\frac{1}{t}$; la qual cosa non poteva non essere; poichè egli, ricorrendo a concetti della meccanica, applicava il principio che lo spazio percorso da un corpo, il quale si muove con moto uniforme, è eguale alla velocità moltiplicata per il tempo:

$$S = v \, . \, t;$$

donde

$$v = \frac{S}{t};$$

e facendo S eguale all'unità:

$$v = \frac{1}{t}.$$

Noi teniamo fermo ad ogni modo alla formola ordinaria o più intelligibile, che dà la *durata* senza far ricorso al concetto di spazio, che davvero non può entrare se non per traslato nei calcoli riguardanti la vita media di una popolazione, la convivenza media dei coniugi, la scadenza media di cambiali, la giacenza media di depositi, ecc.

F) *Rapporti di ripetizione.* — Chiamiamo così i rapporti che si stabiliscono tra il movimento numerico di un fenomeno statistico, il

---

infruttiferi (conti correnti). Non bisogna dimenticare che per disposizione adottata comunemente dagli Istituti di credito l'interesse sui depositi ordinari si calcola per decine o per quindicine *intere* di giorni, a cominciare dal 1° del mese, e che in questi limiti conviene senz'altro al commerciante la forma spiccia del conto corrente. Vi ha dunque margine abbastanza largo, perchè la giacenza delle somme in conto corrente, che varia da 2 a 3 giorni nei paesi industrialmente più evoluti, salga a 10 o 15 in quelli nei quali per le ragioni dette le operazioni d'incassi e di pagamenti hanno carattere di minor continuità. E non v'è insomma bisogno di supporre un rapporto diretto tra lo stato delle finanze di questi secondi paesi e la velocità di circolazione della moneta traverso i conti correnti.

quale può ripetersi più volte sotto le stesse forme, e il movimento del medesimo sotto una forma diversa, definitiva, non ripetibile.

Si muore una volta sola, lasciando una eredità, ma si può essere stati eredi più volte in vita, or come figlio, or come fratello, nipote, ecc., or come estraneo; la stessa persona puo aver figurato più volte nelle statistiche dell'emigrazione del suo paese, e nelle statistiche dei rimpatrii, prima di stabilirsi definitivamente all'estero; le nubili possono essersi fidanzate più volte avanti di contrarre le prime nozze; e si può essere stati sposi e vedovi più d'una volta prima di figurare tra i decessi coniugati o vedovi. E via dicendo.

Allora è chiaro che la osservazione statistica del movimento di un fenomeno ripetibile sotto la stessa forma debba dar luogo a risultati superiori a quelli dell'osservazione del fenomeno stesso nella sua forma definitiva e diversa.

L'esempio, che meglio illustra questo genere di rapporti, lo togliamo dalle statistiche delle successioni ereditarie. In Italia si contano annualmente 150.000 defunti possessori di beni, che vengono ripartiti fra 560.000 eredi. Fra 35 o 40 anni, allorchè gli eredi d'oggi passeranno in generale a miglior vita e ritrasmetteranno quei medesimi beni ed altri ancora alla successiva generazione, è probabile che le statistiche registreranno un 195 o 200 mila defunti lascianti eredità testamentarie o legittime. Questo numero è calcolato all'ingrosso in base allo incremento medio attuale della popolazione italiana.

Esercizi finanziari	Numero delle successioni	Numero degli eredi e legatari
1889-1890	136.276	559.936
1890-1891	138.408	575.513
. . . .	. . .	. . .
1892-1893	156.551	533.021
1893-1894	148.450	583.193
1894-1895	171.024	566.134
. . . .	. . .	. . .
1900-1901	165.913	577.692
1901-1902	152.244	553.070
1902-1903	139.551	521.995

NB. — Per gli esercizi 1891-1892 e 1895-1900 mancano i dati relativi agli eredi e legatari.

Ora si domanda: come mai i 560 mila eredi attuali non compariranno nelle future statistiche annuali (in media fra 35 o 40 anni) come 560 mila autori di successioni, ma come 195 o 200 mila soltanto? La spiegazione non è da ricercarsi nel fatto che molti eredi di minime quote, specie mobiliari, che essi consumano in vita mentre non hanno altro patrimonio proprio, non possono figurare poi come autori di successioni; perchè questo fatto sarebbe compensato dall'altro di coloro che, non avendo mai fatta una eredità in vita loro, si son formati un patrimonio proprio e morendo l'hanno trasmesso. Nemmeno è da

ricercarsi la spiegazione nel fatto che la proprietà, specialmente l'immobiliare la quale non isfugge al fisco, tenda ad accentrarsi in poche mani. La cosa dovrebb'essere dimostrata; d'altronde la grande differenza numerica tra autori di successioni ed eredi indicherebbe un accentramento d'incredibile rapidità. No, la vera ragione di cotesta differenza è che si può essere eredi più volte in vita, mentre si lascia eredità una volta sola. Ogni anno abbiamo 150 mila persone in media di cui si apre la successione legittima o testamentaria; ma esse sono, tutte e ciascuna, diverse da quelle che morirono in qualsiasi degli anni antecedenti e che moriranno in qualsiasi degli anni avvenire; invece i 560 mila eredi d'oggi non sono tutte persone diverse dagli eredi registrati nelle statistiche passate o che si registreranno nelle venture, lo stesso individuo potendo aver ereditato più volte in vita sua, or come figlio, or come nipote, cugino, estraneo, ecc. Supponiamo per intenderci meglio, che ognuno degli attuali deceduti, lascianti eredità, abbia in vita sua ereditato da altri un paio di volte ed abbia figurato quindi due volte nelle statistiche degli eredi di cinque, dieci, venti o più anni fa; allora è chiaro che alle schiere attuali di deceduti, composte di 150 mila individui annualmente, debbono aver fatto riscontro schiere passate di eredi, due volte più numerose, cioè composte in media annua di 300 mila persone. Se invece ognuno degli attuali deceduti aveva ereditato *tre* volte in vita sua, cioè aveva figurato tre volte nelle passate statistiche degli eredi, evidentemente alle schiere attuali di deceduti, composte di 150 mila persone all'anno, debbono aver corrisposto schiere di eredi composte di 450 mila individui (450.000 = 3 × 150.000). E così via. Si conclude dunque, che se alle attuali schiere di eredi forti di 560.000 individui all'anno corrisponderanno in capo ad una generazione schiere di 195 o 200 mila autori di successioni testate o intestate, ognuno di questi avrà figurato come erede nelle statistiche attuali tante volte quanto il 195.000 o il 200.000 sta in 560.000, ossia 2,8 volte in media.

Sicchè, prescindendo da cause secondarie, che influiscono sul rapporto, è lecito affermare che in Italia, *tra le classi abbienti*, la qualità di erede si ripete, vita natural durante, da due volte e mezza a tre nella stessa persona.

§ **3.** Dei rapporti, che possono assumersi come espressione della *probabilità* di un fenomeno, sarà parola nel capitolo dedicato alla « teoria delle probabilità ». Dei rapporti, che si stabiliscono fra dati non primitivi, ma derivati, e di cui ci gioviamo come *indici di situazioni complesse*, discorreremo nel capitolo concernente i « numeri-indici ». E del pari devesi rimandare al capitolo sull'interpolazione delle « serie » un cenno sui rapporti, i cui termini non sono quantità semplici o uniche, ma serie di quantità statistiche. Perciò passiamo senz'altro all'argomento delle *perequazioni*.

C) *Perequazioni.*

Sommario: § 1. Perequazione per medie aritmetiche semplici. — § 2. Perequazione per medie di medie, ossia per medie aritmetiche ponderate; critica del metodo.

§ **1.** Allorquando i dati di una serie o di una seriazione si succedono con una irregolarità, che v'è ragione d'imputare soprattutto al modo imperfetto della loro rilevazione, conviene « perequarli » ossia livellarne le ineguaglianze e restituire alla serie la sua naturale continuità mediante la sostituzione di valori opportunamente calcolati. All'uopo servono le medie aritmetiche semplici o ponderate. Abbiamo già visto (pag. 82) che la ripartizione dei censiti, anno per anno d'età, presenta agglomerazioni singolari in corrispondenza delle età rotonde (terminanti per 0 o per 5) e deficienze negli anni dispari (eccetto il 5); la qual cosa si spiega coll'abitudine, che è in molti ignari o immemori della loro vera età, di indicare questa in maniera approssimativa, arrotondandola in decine o mezze decine. Ora, siccome l'ipotesi più ragionevole è che la serie dei censiti diminuisca gradatamente col crescere dell'età, il livellamento delle ineguaglianze e discontinuità delle cifre greggie del censimento si può attuare col sostituire ad ogni termine dato la media aritmetica dei tre termini, oppure dei cinque, dei sette termini, ecc., di cui esso è centrale. Qui si opererebbe per medie aritmetiche semplici. Sostituendo poi ad ogni termine dato le medie delle medie si opera, come vedremo, per medie ponderate. Esemplifichiamo il primo metodo, valendoci dei dati riportati già in nota a pag. 82 (V. prospetto a pagina seguente).

Dunque alla classe d'età, per es., da 31 a 32 anni, rilevata in 33.516 individui noi possiamo sostituire come valore più verosimile la media aritmetica delle classi 30-31, 31-32 e 32-33, che è eguale a 36.354, ovvero la media aritmetica delle cinque classi 29-30, 30-31, 31-32, 32-33, 33-34, che è eguale a 34.931; ovvero la media aritmetica delle sette classi, di cui quella da perequare forma il termine di mezzo, media che il nostro prospetto indica in 34.429. E così via.

Prendendo le differenze fra un termine e l'altro, così della serie greggia, come delle serie perequate, è facile verificare che in queste le discontinuità e ineguaglianze dei dati originarii sono molto attenuate e la condizione di un regolare decrescere delle classi, anno per anno d'età, è abbastanza bene soddisfatta.

Non bisogna però farsi molte illusioni sulla bontà del metodo. Anzitutto è affatto arbitrario il numero di termini, su cui si operano le medie, nè l'arbitrio viene in qualche modo disciplinato, come lo è nei processi d'interpolazione ove si prestabilisce una condizione da osservarsi da tutti i calcolatori: la condizione di un minimo nella

somma dei quadrati degli scostamenti fra la serie data e la serie teorica. In secondo luogo la somma dei termini perequati non riproduce esattamente quella dei dati. Il metodo poi non si può applicare ai primi termini della serie. Infatti, ad es., per i bambini da 0 a 1 anno non esiste un'età antecedente, che insieme a quella data e alla successiva possa formare materia di una media di tre termini; e a maggior ragione dicasi se si tratta di medie di 5, 7 e più termini (1). Infine la perequazione nella forma anzidetta non può applicarsi ai tratti in cui la serie sia crescente o decrescente, in ragion geometrica; perchè il termine di mezzo verrebbe allora sistematicamente ad ingrandirsi o ad impicciolirsi.

CAPOLUOGHI DI PROVINCIA.

MASCHI CENSITI NEL 1881		CIFRE PEREQUATE		
Età	Dati greggi	Per medie di tre termini	Per medie di cinque termini	Per medie di sette termini
28-29	36.025	...	...	...
29-30	32.626	37.297	...	...
30-31	43.241	36.461	35.542	...
31-32	33.516	36.354	34.931	34.429
32-33	32.304	32.929	34.471	34.171
33-34	32.966	31.867	32.666	34.206
34-35	30.330	32.504	32.537	32.358
35-36	34.216	32.472	32.137	32.016
36-37	32.871	32.463	31.768	31.269
37-38	30.301	31.431	31.117	32.500
38-39	31.121	29.499	32.591	32.492
39-40	27.075	33.260	32.071	31.950
40-41	41.585	32.977	32.095	30.799
41-42	30.272	34.094	30.834	30.185
42-43	30.424	28.503	30.620	...
43-44	24.814	27.081	...	...
44-45	26.005	...	...	...

Ad ogni modo, quando si voglia fare una perequazione aritmetica, in vista della maggior facilità e speditezza di calcolo che essa presenta

(1) Questo inconveniente manca però nelle serie cicliche e periodiche. Così per perequare le temperature giornaliere dell'annata, a cominciare dal 1° gennaio, posto che si proceda per medie di 11 termini, non avremo altro che da sostituire al dato del 1° gennaio la media aritmetica delle temperature registrate giornalmente dal 27 dicembre al 6 gennaio; al dato del 2 gennaio, la media delle temperature dal 28 al 7; e così al dato del 3 gennaio, la

in confronto dell'interpolazione, devesi bene avvertire che gli errori della serie greggia non abbiano carattere costante. Se, come nel caso delle denunzie d'età dei censiti, essi cadono sistematicamente in corrispondenza di certi valori della variabile, converrà fare un'analisi comparativa dei diversi tratti di serie inquinati da errori e valutare di questi, pur con qualche latitudine, la quantità, direzione ed estensione di effetto. Già abbiamo dimostrato che le agglomerazioni attorno alle età terminanti per zero si producono soprattutto a scapito delle due età immediatamente contigue e poi, in grado minore, delle altre contigue a queste; sì che può ritenersi l'attrazione dell'età rotonda si estenda fino a due anni e mezzo prima e due anni e mezzo dopo; laddove l'attrazione dell'età terminante per 5 non sembra arrivare al di là di un anno e mezzo prima e di uno e mezzo dopo. Appariva dunque buon partito sostituire alla classe dell'età in zero la sesta parte della somma formata con questa medesima classe e le due e mezza antecedenti e susseguenti; e alla classe dell'età in 5, la quarta parte della somma formata con essa classe, le due contigue e mezza di ciascuna delle altre due contigue a queste. Determinato così il valore più probabile delle due classi in 0 e in 5, che sono le maggiormente inquinate d'errori, non restava che calcolare le intermedie con un facile processo d'interpolazione.

La necessità di eliminare, sia pure *grosso modo*, preventivamente gli errori costanti, s'impone per ciò, che la perequazione aritmetica applicata per medie semplici, pur attenuando nell'insieme gli errori, tende a riversarli su punti determinati. Chi analizzasse un poco le serie addietro riferite, troverebbe che ivi la perequazione per medie di 3 termini avvantaggia (o danneggia) sistematicamente le classi contigue a quella che nella serie originaria presenta il maggiore agglomeramento (o la maggior deficienza) di censiti; troverebbe pure che la perequazione per medie di 5 termini avvantaggia o danneggia sistematicamente le classi che di due anni precedono o susseguono a quella da perequare, ecc. Della qual cosa si potrebbe anche dare una dimostrazione generale.

Valga un secondo esempio. S'è visto che in Sicilia un certo numero di maschi nati a fin d'anno sogliono essere denunziati come nati nel gennaio dell'anno successivo. Il modo più ovvio di correggere l'errore, che ne deriva nel calcolo dei rapporti delle nascite maschili alle femminili, è di attribuire tanto al dicembre quanto al gennaio un rap-

---

media dal 29 all'8, ecc. Or bene, quand'anche non si avessero a disposizione le registrazioni dell'anno antecedente o del successivo, non ne sarebbe impedito il perequare i dati dei primi e degli ultimi giorni dell'anno considerato, potendosi utilizzare rispettivamente le ultime e le prime registrazioni di questo stesso anno, come se appartenessero le une al dicembre dell'anno antecedente, le altre al gennaio dell'anno successivo.

porto medio fra quelli risultanti dai dati greggi. Se invece perequassimo per medie di tre termini, evidentemente il novembre ed il febbraio risulterebbero in modo sistematico l'uno danneggiato, l'altro avvantaggiato. Infatti al dato del novembre dovrebbesi sostituire la media aritmetica dei dati dell'ottobre, del novembre stesso e del dicembre; e la scarsità eccezionale di maschi denunziati nel dicembre andrebbe dunque, sebbene attenuata, a danno del termine da perequare, cioè del novembre. Così al dato del febbraio dovrebbesi sostituire la media aritmetica di quelli del gennaio, del febbraio stesso e del marzo; e l'abbondanza eccezionale di maschi denunziati nel gennaio andrebbe, sebbene attenuata, a favore del termine da perequare, il febbraio. Queste considerazioni hanno una portata men piccola di quella che a prima giunta potrebbe ritenersi, perchè appunto nella maggior parte dei casi l'analisi riesce a rivelare negli errori un qualche aspetto costante; ed esse, unite alle altre che già facemmo, tolgono quasi ogni valore a questo metodo e in suo confronto ingrandiscono il valore dei processi interpolatorii.

§ **2.** Una variante della perequazione per medie aritmetiche si ha replicando il processo stesso sulle medie già ottenute. Diamo qui un esempio.

Età	Maschi censiti nei capoluoghi di provincia nel 1881	Perequazione per medie di tre termini	Perequazione per medie di tre medie
28-29	36.025	. . .	. . .
29-30	32.626	37.297	. . .
30-31	43.241	36.461	36.704
31-32	33.516	36.354	35.248
32-33	32.304	32.929	33.717
33-34	32.966	31.867	32.433
34-35	30.330	32.504	32.281
35-36	34.216	32.472	ecc.
ecc.	ecc.	ecc.	

Alla cifra greggia di 43.241 censiti in età da 30 a 31 noi possiamo sostituire quella di 36.461, che risulta dalla media aritmetica delle classi 29-30, 30-31, 31-32, ossia dall'espressione:

$$\frac{32.626 + 43.241 + 33.516}{3};$$

ma possiamo anche sostituirle la cifra 36.704, che risulta dalla media delle tre medie 37.297, 36.461, 36.354, pure corrispondenti alle età 29-30, 30-31, 31-32. Questo artificio, come è facile dimostrare, si riduce a fare sui numeri originarii una perequazione per medie ponderate, in cui il dato da perequare figura con un peso o coefficiente d'impor-

tanza di 3, i due dati contigui con un coefficiente di 2, e gli altri due contigui a questi con un coefficiente di 1. In altri termini, è facile verificare che sussistono le eguaglianze:

$$36.704 = \frac{37.297 + 36.461 + 36.354}{3} =$$

$$= \frac{(1 \times 36.025) + (2 \times 32.626) + (3 \times 43.241) + (2 \times 33.516) + (1 \times 32.304)}{9}$$

Diamo ad ogni modo la dimostrazione generale. Siano

$$a_1, a_2, a_3, a_4, a_5 \dots$$

i numeri greggi di una rilevazione. Al termine a_2 possiamo, perequando, sostituire la media aritmetica di a_1, a_2, a_3; e così al termine a_3, la media aritmetica di a_2, a_3, a_4; e al termine a_4 la media aritmetica di a_3, a_4, a_5. Se applichiamo due volte il metodo, sostituendo ad a_3 la media delle tre medie già trovate corrispondenti ai termini a_2, a_3, a_4, avremo:

$$a_1$$

$$a_2 = \frac{a_1 + a_2 + a_3}{3}$$

$$a_3 = \frac{a_2 + a_3 + a_4}{3} = \frac{\frac{a_1 + a_2 + a_3}{3} + \frac{a_2 + a_3 + a_4}{3} + \frac{a_3 + a_4 + a_5}{3}}{3}$$

$$a_4 = \frac{a_3 + a_4 + a_5}{3}$$

$$a_5$$

$$= \frac{a_1 + 2a_2 + 3a_3 + 2a_4 + a_5}{9}$$

Naturalmente il segno di *eguaglianza* deve intendersi come segno di *sostituzione*. Dunque il dato greggio a_3 può essere sostituito dall'espressione:

$$\frac{a_1 + 2a_2 + 3a_3 + 2a_4 + a_5}{9},$$

che non è altro se non una media ponderata nella maniera anzidetta.

Se avessimo fatta la prima perequazione per medie di 5 termini e la seconda per medie di 5 medie, saremmo giunti alla formola (a cominciare dal termine a_5):

$$a_5 = \frac{a_1 + 2a_2 + 3a_3 + 4a_4 + 5a_5 + 4a_6 + 3a_7 + 2a_8 + a_9}{25}$$

che è ancora una media ponderata, in cui al termine da perequare si dà un coefficiente di 5, ai due contigui un coefficiente di 4, agli altri due contigui a questi un coefficiente di 3, e così via.

In occasione del censimento del 1881 si è proceduto per i dati di tutto il regno rilevati in classi quinquennali perequando mediante medie di 3 medie, ossia colla formola:

$$a_n = \frac{a_{n-2} + 2a_{n-1} + 3a_n + 2a_{n+1} + a_{n+2}}{9}.$$

Classi d'età	Numero greggio dei censiti	Numeri perequati
25-30	2.127.602	. . .
30-35	2.056.254	. . .
35-40	1.776.496	1.837.928
40-45	1.795.475	1.690.967
45-50	1.380.802	1.500.346
50-55	1.461.432	. . .
55-60	1.070.397	. . .
. . .	. . .	. . .

Siffatto lavoro di perequazione era consigliabile per eliminare in qualche modo le irregolarità della serie prodotte dalle erronee denunzie d'età, le quali irregolarità si fanno manifeste anche nella classificazione per gruppi di cinque in cinque anni. Le età rotonde, essendo quelle da 30 a 31, da 40 a 41, da 50 a 51, ecc., vengono colle loro agglomerazioni di censiti ad ingrossare le classi quinquennali da 30 a 35, da 40 a 45, da 50 a 55, ecc., a scapito delle immediatamente precedenti. Ed infatti, esaminando i dati greggi del censimento del 1881 si trova che il gruppo da 40 a 45 è più numeroso di quello da 35 a 40, e similmente che il gruppo da 50 a 55 è più numeroso di quello da 45 a 50, ecc., mentre di regola deve aversi l'opposto, perchè data una popolazione, come l'italiana, da lungo tempo in aumento, le classi più vecchie derivano da generazioni meno numerose di quelle da cui derivano le più giovani e, oltre a ciò, sono state esposte a morire cinque anni di più.

Alla perequazione per medie di medie si possono fare pressochè le stesse critiche, già formulate a proposito di quella per medie semplici. Il metodo è arbitrario e all'arbitrio non è posta nessuna speciale disciplina; la somma dei termini perequati non può riprodurre con esattezza quella dei termini corrispondenti, sicchè devesi ricorrere ad un aggiustamento, pure arbitrario, delle cifre; i primi termini della serie restano esclusi dalla perequazione, e gli errori, in quanto abbiano un carattere di costanza, vengono, sebbene attenuati, a riversarsi su punti determinati della serie. Per queste ragioni non crediamo di insistere più a lungo sull'argomento (1), molto più che il fine della perequazione appare meglio raggiungibile col metodo dell'interpolazione.

---

(1) Per la trattazione matematica rimandiamo allo scritto di L. PEROZZO, *Nota sui principali metodi di perequazione usati nelle scienze d'osservazione*, negli *Annali di Statistica*, serie III, vol. XVI, pag. 61 e seg., nonchè ai lavori ivi citati, specialmente quelli del WITTSTEIN e dello SCHIAPARELLI.

TITOLO II.

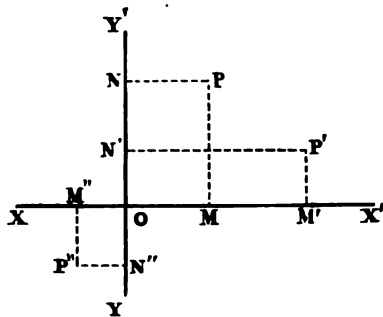
PROCEDIMENTI GEOMETRICI E GRAFICI (1).

A) *Diagrammi.*

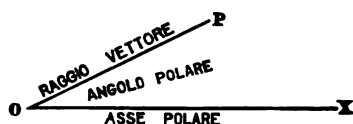
Sommario: § 1. Sistema *cartesiano* e sistema *polare*. — § 2. Rappresentazione di funzioni algebriche e trigonometriche. — § 3. Caso di seriazioni a classi inegualmente estese. — § 4. Scale di misura espresse in termini qualitativi. — § 5. Diagrammi a base circolare. — § 6. Diagrammi a scala logaritmica. — § 7. Rappresentazione di fenomeni a due variabili. — § 8. Rappresentazione di fenomeni a tre variabili; stereogrammi.

§ 1. Le quantità statistiche rilevate e i rapporti su di esse calcolati si possono rendere di ancor più facile comprensione coi metodi geometrici e grafici. Le forme principali di rappresentazione, cui questi dànno luogo, sono i *diagrammi* e i *cartogrammi*.

La posizione di un punto nel piano si determina, come insegna la geometria analitica, riferendo questo punto ad elementi supposti fissi nel piano medesimo, cioè ad un *sistema di coordinate*. Nel sistema detto « cartesiano » (dal nome di Descartes, italianamente Cartesio, che primo lo introdusse nel 1637) gli elementi fissi sono due rette, di cui l'una XX′ dicesi *asse delle ascisse*, l'altra YY′ dicesi *asse delle ordinate*, che incontra la prima nel punto O (*origine delle coordinate*). L'angolo formato dai due assi può essere qualunque; se è retto, il sistema prende il nome di *rettangolare* od *ortogonale*. Per le rappresentazioni di fenomeni statistici il sistema ortogonale è d'uso generalissimo.

Così nella figura qui a fianco la posizione del punto P è individuata nel piano in cui giacciono le rette XX′ e YY′, quando, condotte da esso punto le perpendicolari PM, PN ai due assi, si conoscono in *grandezza* e in *senso* i segmenti OM e ON, che sono rispettivamente eguali a PN e PM. Si conviene di dare ad OM segno positivo o negativo, secondo che M si trova a destra o a sinistra di O, e di dare

(1) I procedimenti geometrici sono anch'essi « grafici »; tuttavia noi preferiamo applicare la denominazione di « grafici » in senso ristretto alle forme di rappresentazione, in cui uno degli elementi dimensivi è fornito dalle gradazioni di uno o più colori oppure dall'addensamento più o meno notevole di punti in un dato spazio, come spesso vediamo nelle carte statistico-geografiche a tinte graduate e in altre forme di cartogrammi.

ad ON segno positivo o negativo secondo che N si trova al di sopra o al di sotto di O. Ad esempio, la posizione del punto P″, che figura a sinistra e in basso di O, sarebbe individuata dai due segmenti di segno negativo OM″, ON″, rispettivamente eguali alle perpendicolari P″N″ e P″M″. Per la determinazione delle grandezze dei segmenti importa prestabilire una *unità di misura* (il centimetro, il millimetro), ossia prestabilire una *scala graduata* su ciascuno dei due assi coordinati.

La posizione di un punto nel piano può essere anche determinata, quando si conosca la sua distanza da un punto fisso O (*polo*) e l'angolo che il *raggio vettore* (cioè la retta che unisce il punto dato con O)

forma con una retta fissa OX passante per il polo (*asse polare*). Ma i diagrammi secondo questo sistema si prestano men bene agli usi della statistica e costituiscono rappresentazioni di puro lusso. Noi ci atterremo al sistema più semplice e ormai popolare, che è il cartesiano ortogonale.

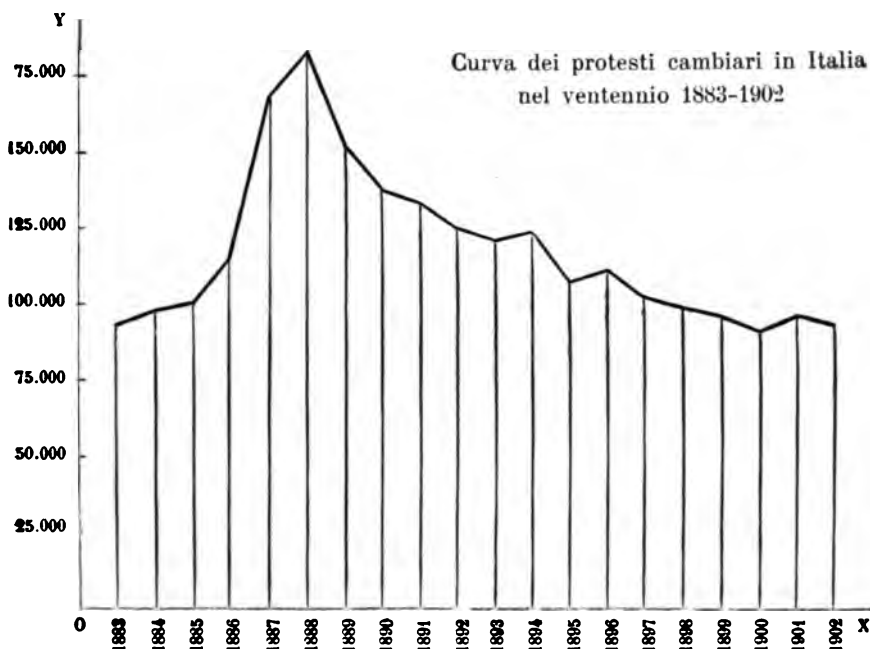
Sorvolando anche sul problema della trasformazione delle coordinate, intorno al quale lo studioso potrà consultare qualsiasi buon trattato elementare di geometria analitica (1), vediamo come si può tradurre in diagramma una successione di quantità statistiche per rapporto al tempo o ad altra variabile.

Abbiasi la serie dei protesti cambiarii verificatisi in Italia dal 1883 al 1902.

Anni	N° dei protesti	Anni	N° dei protesti
1883	94.168	1893	121.395
1884	98.766	1894	123.831
1885	102.524	1895	109.085
1886	115.985	1896	112.345
1887	170.973	1897	104.290
1888	184.704	1898	99.634
1889	154.498	1899	95.829
1890	139.263	1900	91.988
1891	134.062	1901	96.353
1892	125.962	1902	93.335

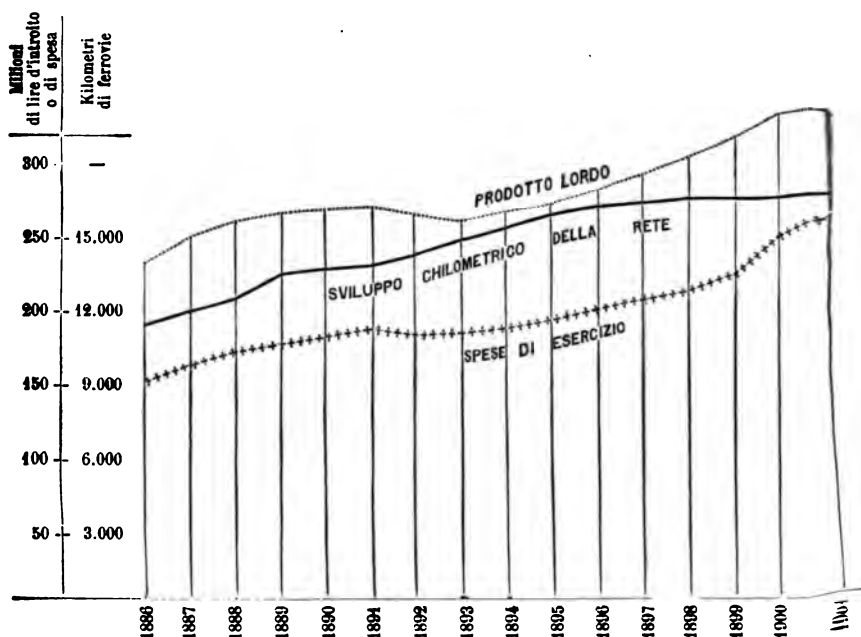
Sulla OX riportiamo la serie dei tempi e poichè questi procedono per intervalli eguali (anni), così prenderemo sulla OX, a partire da O una successione di segmenti eguali, poniamo, a mezzo centimetro ciascuno. Alla prima divisione scriverò 1883, oppure anno 1°; alla

(1) Veggasi pure il *Corso speciale di matematiche* di Mineo Chini, che contiene numerose applicazioni ad uso principalmente dei chimici e dei naturalisti. Livorno, R. Giusti, 1904.

seconda, 1884, oppure anno 2°, e così via. Sull'asse O Y delle ordinate fisseremo una scala di misura per la frequenza dei protesti, prendendo, ad es., per la cifra di 25.000 protesti, un segmento di un centimetro e fissando quindi i punti 50.000, 75.000, 100.000, ecc.: a distanze di 2, 3, 4, ecc. centimetri dall'origine 0:

Questo diagramma dovrebbe dirsi cartesiano ortogonale *a ordinate congiunte*. Infatti gli estremi delle ordinate ivi son congiunti da una linea « spezzata » che impropriamente si conviene di chiamare « curva » per agevolare la continuità della visione; mentre se le ordinate fossero rimaste disgiunte, l'occhio non potrebbe seguire l'andamento del fenomeno rappresentato, se non procedendo, per così dire, a salti da un'ordinata all'altra.

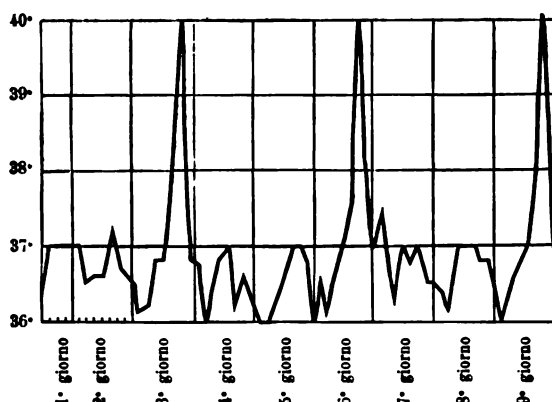
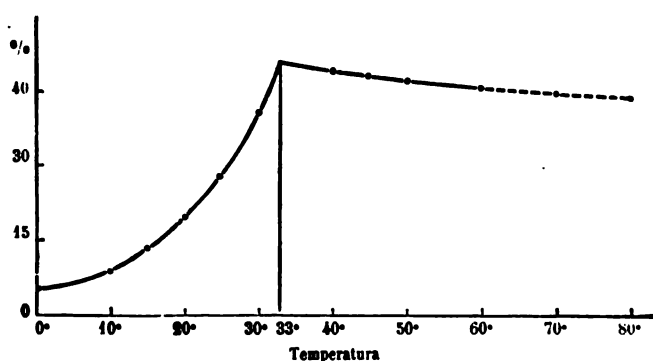
L'occhio apprezza più facilmente le differenze assolute che i rapporti delle grandezze graficamente rappresentate; apprezza bene anche gli angoli che le rette congiungenti gli estremi delle ordinate fanno colle ordinate medesime. L'allontanarsi di una di queste congiungenti dall'asse delle ascisse indica che il fenomeno cresce di frequenza, l'avvicinarsi ossia il discendere di quella verso l'asse delle ascisse indica l'opposto.

Su uno stesso diagramma possono essere rappresentati due o più fenomeni riferiti alla stessa variabile, anche se per essi debbansi adottare scale di misura diverse. Così, nel diagramma che qui facciamo seguire, l'asse delle ordinate porta due scale, l'una concernente il numero

di chilometri di ferrovie aperte all'esercizio in Italia, l'altra l'ammontare in milioni di lire del prodotto lordo e delle spese d'esercizio. Il tracciato a linea continua figura lo sviluppo chilometrico delle reti; gli altri due a linea punteggiata e tratteggiata figurano rispettivamente gli introiti e le spese.

Per non ingenerare confusione conviene, per quanto è possibile, stabilire le scale di misura in maniera che i diversi tracciati non si intersechino mai o s'intersechino il minor numero di volte.

Si ottiene economia di spazio facendo incominciare la scala delle ordinate non dallo zero, ma da una cifra rotonda qualunque, purchè inferiore al valor minimo, cui discende la serie da rappresentare. Per fenomeni, che oscillano intorno ad un valore, il quale costituisca in certo modo lo stato normale, giova prendere come asse delle ascisse l'ascissa corrispondente a cotesto valor normale. Nel primo dei diagrammi che seguono, in cui è figurato l'andamento della temperatura del corpo nella febbre detta *quartana semplice regolare*, abbiamo appunto preso come asse delle ascisse l'ascissa corrispondente a 37 centigradi, temperatura normale del corpo sano. Ivi la serie delle ordinate comincia dal mezzodì del primo giorno, alla temperatura di 36°,5.

Il vantaggio immediato della rappresentazione grafica è quello di permettere allo sguardo di abbracciare d'un colpo l'andamento del fenomeno nelle sue ondulazioni caratteristiche. Da ciò l'uso omai popolare del metodo in demografia, in meteorologia, in chimica, in medicina, ecc.

Curva della temperatura nella *quartana semplice regolare.*

Quest'altro diagramma ci dà la curva di solubilità del *solfato di sodio* idrato, cioè rappresenta quante parti in peso di questa sostanza sono sciolte in 100 parti in peso d'acqua, col variare della temperatura

Curva di solubilità del *solfato di sodio* (sale di Glauber).

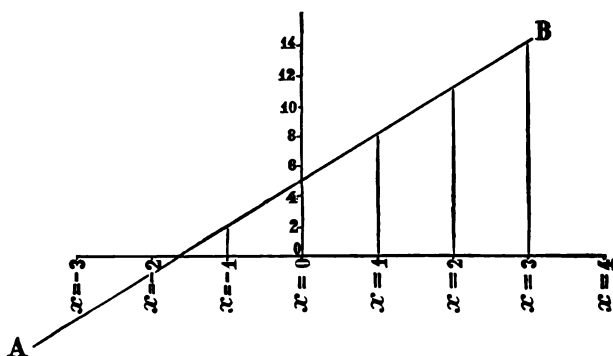
dell'acqua stessa. La curva si segnala tosto per una cuspide assai caratteristica in corrispondenza della temperatura di 33°. La cuspide si spiega col fatto che fino a 33° la forma stabile del solfato di sodio è $Na_2SO_4.10H_2O$ (sale *idrato*); mentre a temperature superiori la forma stabile è Na_2SO_4 (sale *anidro*), avente una solubilità diversa da quella dell'idrato, tale anzi che diminuisce col crescere della temperatura.

§ 2. Per mezzo di diagrammi si possono rappresentare le funzioni algebriche intere, le funzioni trigonometriche ed altre, il cui impiego è frequente in statistica. Cominciamo dal caso più semplice dell'equazione di una retta.

Sia la funzione: $y = 3x + 5$.

Se facciamo assumere alla variabile x i successivi valori 0, 1, 2, 3, ecc., è chiaro che anche y assumerà certi altri valori. Così per $x=0$, sarà $y=(3\times 0)+5=5$; per $x=1$, sarà $y=(3\times 1)+5=8$; per $x=2$, sarà $y=(3\times 2)+5=11$, e via dicendo. Possiamo attribuire ad x anche dei valori negativi; ad es.: per $x=-1$, risulta $y=2$; per $x=-2$, risulta $y=-1$; per $x=-3$, risulta $x=-4$, ecc. Segnando dunque sull'asse delle ascisse i successivi valori di x e su quello delle ordinate la scala dei valori di y, l'equazione $y=3x+5$ verrà così rappresentata dalla retta AB:

TABELLA DEI VALORI

Per	$x = -3$,	$y = -4$
»	$x = -2$,	$y = -1$
»	$x = -1$,	$y = 2$
»	$x = 0$,	$y = 5$
»	$x = 1$,	$y = 8$
»	$x = 2$,	$y = 11$
»	$x = 3$,	$y = 14$
	ecc.	ecc.

Si potrebbe dimostrare in maniera generale che ogni serie aritmetica, cioè procedente per *equidifferenze*, può essere graficamente rappresentata da una retta, algebricamente da una equazione della forma:

$$y=ax+b.$$

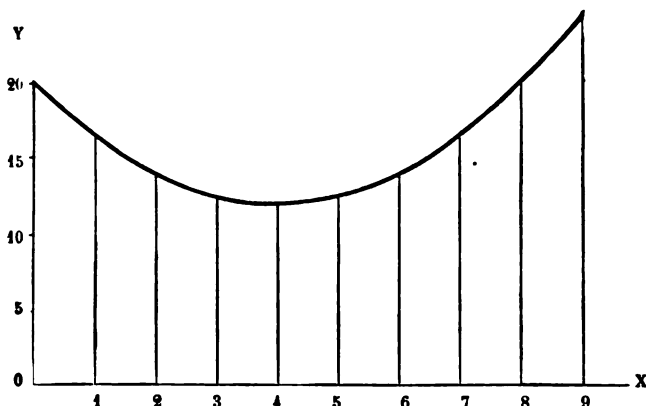
Facciamo ora un caso alquanto meno semplice. Sia la funzione:

$$y=\frac{x^2}{2}-4x+20.$$

Procedendo come poc'anzi, daremo ad x i successivi valori 0, 1, 2, 3, ecc. Per $x=0$, si ha evidentemente $y=20$; per $x=1$, si avrà invece $y=\frac{1}{2}-4+20=16{,}5$; per $x=2$, $y=\frac{4}{2}-8+20=14$ e

così via come dalla tabella che qui riportiamo, accanto alla quale sta la rappresentazione geometrica della funzione data:

TABELLA DI VALORI DELLA FUNZIONE $y = \frac{x^2}{2} - 4x + 20$

Per $x = 0$,	y 20
» $x = 1$,	y 16,5
» $x = 2$,	y 14
» $x = 3$,	y 12,5
» $x = 4$,	y 12
» $x = 5$,	y 12,5
» $x = 6$,	y 14
» $x = 7$,	y 16,5
ecc.	ecc.

La curva che rappresenta l'equazione data è una *parabola ordinaria* o di 2° grado. Si potrebbe dimostrare in maniera generale che una serie, le cui differenze prime costituiscono una progressione aritmetica e le differenze seconde (differenze delle differenze) sono una costante, può essere graficamente rappresentata da una parabola ordinaria, algebricamente da una equazione della forma $y = ax^2 + bx + c$.

Il lettore può adesso, con analogo procedimento, rappresentare geometricamente equazioni di 3°, 4° grado, ecc., ovvero funzioni di questa forma: $\frac{k}{x^m}$ (curve iperboliche, di cui un caso particolare è l'*iperbole equilatera* dall'equazione: $y = \frac{k}{x}$, dove k è una costante qualunque), che pure ricorrono di frequente nei fenomeni demografici, economici, ecc., ed altre ancora. Qui non daremo che un esempio di rappresentazione di una formola trigonometrica.

Abbiamo in Statistica frequenti casi di serie periodiche, cioè di fenomeni, i quali presentano un andamento a massimi e a minimi ripetentisi a intervalli regolari (anno, settimana, giorno). Per queste

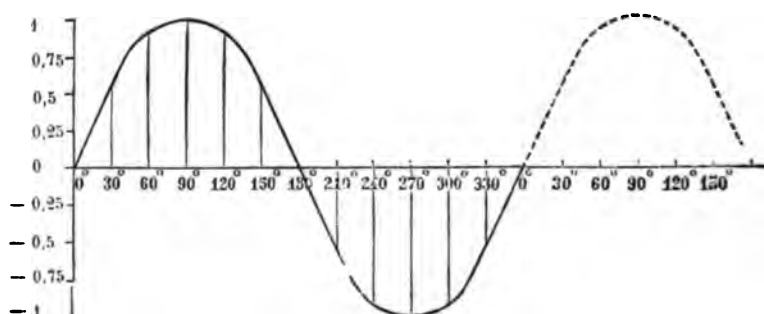
serie la rappresentazione analitica dà luogo a formole, in cui entrano elementi trigonometrici, come il *seno*, il *coseno*, ecc.

È risaputo che il *seno* di un angolo (ossia il rapporto tra la perpendicolare opposta all'angolo e l'ipotenusa o raggio) è zero per un angolo di zero gradi; è eguale all'unità per un angolo di 90°. Col variare dell'angolo da 0° a 90°, il rapporto anzidetto assume valori diversi compresi fra zero e l'unità. Se l'angolo è di 30°, allora la perpendicolare è la metà esatta dell'ipotenusa o raggio (v. fig. a pag. 109, nota), cioè *sen* 30° = 0,5. Se l'angolo è di 45°, la perpendicolare è poco più di $^7/_{10}$ del raggio, ossia *sen* 45° = 0,7 (più esattamente 0,707107...). Per un angolo di 80°, equivarrebbe a circa 94 % del raggio: *sen* 80° = 0,939693... E così via. Orbene i valori del *seno*, che per angoli da 0° a 90° crescono da zero all'unità, diminuiscono simmetricamente dall'unità a zero per angoli crescenti da 90° a 180°; poi diminuiscono ancora da 0 a —1 nel 3° quadrante (cioè per gli angoli crescenti da 180° a 270°); infine risalgono da —1 a 0 nel 4° quadrante (ossia per angoli crescenti da 270° a 360°) per ricominciare daccapo allorchè il raggio vettore ripassa sul primo quadrante.

Sia da rappresentare la funzione: $y = sen\ \varphi$.

Diamo all'angolo φ (la variabile) le successive ampiezze di 0°, di 30°, di 60°, di 90°, di 120°, ecc.; allora i valori di *sen* φ diventeranno rispettivamente = 0; 0,5; 0,866; 1; 0,866, ecc., come dalla tabella qui sotto.

φ	Valori di *sen* φ	φ	Valori di *sen* φ
0°	0	240°	— 0,866
30°	0,5	270°	— 1
60°	0,866	300°	— 0,866
90°	1	330°	— 0,5
120°	0,866	360° = 0° . . .	0
150°	0,5	390° = 30° . . .	0,5
180°	0	420° = 60° . . .	0,866
210°	— 0,5	ecc.	ecc.

Riportando poi sull'asse delle ascisse i valori di φ in gradi e su quello delle ordinate i valori di *sen* φ, otterremo la curva periodica ben nota sotto il nome di *sinusoide*.

Data adunque una equazione, il problema di descriverne la curva non presenta difficoltà di sorta. Ne presenta invece il problema inverso, quello cioè pel quale, data la curva, si tratta di stabilire la sua equazione esatta o approssimata. Ma di esso sarà parola nel capitolo dell'interpolazione, dove peraltro ci limiteremo ai procedimenti più facili e d'interesse immediato per la Statistica.

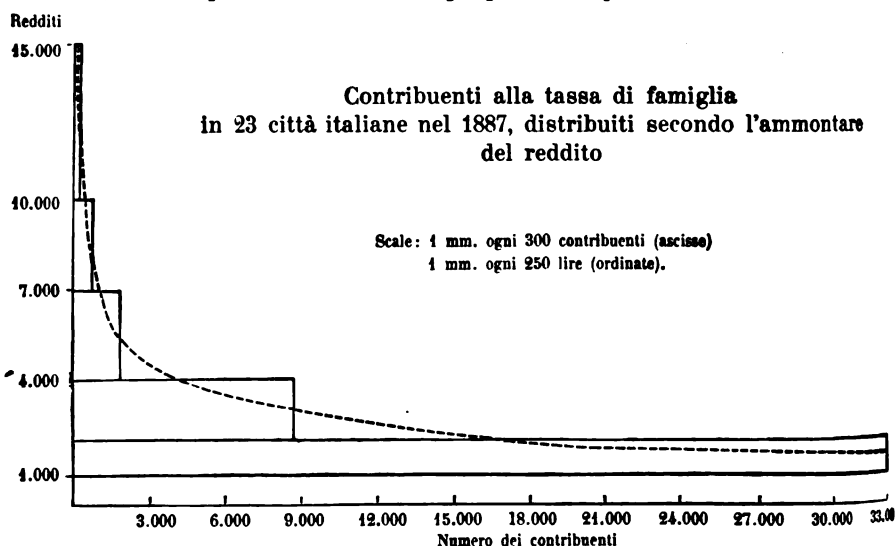
§ 3. Alle volte la serie o seriazione da rappresentare procede per classi o gruppi di elementi riferiti a intervalli inegualmente estesi dei valori della variabile, come allorquando una popolazione è distinta nelle prime età per classi annuali e nelle età successive per classi triennali o quinquennali o decennali; ovvero quando un gruppo di contribuenti è ripartito per categorie, che procedono, mettiamo, prima di 1000 in 1000 lire di reddito imponibile, poi di 2000 in 2000, di 5000 in 5000, di 10.000 in 10.000, ecc. Fermiamoci appunto a questo esempio.

CONTRIBUENTI ALLA TASSA DI FAMIGLIA IN 23 CITTÀ ITALIANE NEL 1887.

Reddito	N° dei contribuenti
Da 1.000 a 2.000 lire	32.518
» 2.000 a 4.000 »	17.202
» 4.000 a 7.000 »	5.502
» 7.000 a 10.000 »	1.867
» 10.000 a 15.000 »	1.087
» 15.000 a 25.000 »	665
Oltre 25.000 lire	645
Totale . . .	59.486

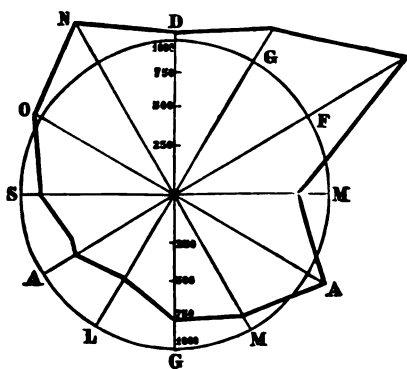
Gli è chiaro che la rappresentazione grafica dovrà tener conto della diversa estensione delle categorie. Fissiamo dunque sull'asse delle x (ascisse) la scala di frequenza dei contribuenti, su quello delle y (ordinate) i limiti delle categorie di reddito. Il primo gruppo di contribuenti sarà rappresentato dal rettangolo avente per base il segmento corrispondente alla cifra 32.518 sull'asse delle ascisse, e per altezza il segmento da 1000 a 2000. Il secondo gruppo comprendendo 17.202 individui nei limiti da 2000 a 4000 lire, cioè in limiti due volte più ampi di quelli del gruppo precedente, si può considerare, provvisoriamente, come composto di 8601 $\left(= \frac{17.202}{2}\right)$ contribuenti con reddito da 2000 a 3000 lire ed altri 8601 con reddito da 3000 a 4000. Perciò il rettangolo, che figura tutto il secondo gruppo, dovrà avere per base un segmento corrispondente alla cifra 8601 sull'asse delle ascisse, e per altezza un segmento preso sull'asse delle ordinate di lunghezza doppia di quella del primo gruppo. Similmente la terza classe di contribuenti nei limiti da 4000 a 7000 lire di reddito, che novera 5502 persone, potrà considerarsi composta di 1834 $\left(= \frac{5502}{3}\right)$ persone aventi da 4000

a 5000 lire, di altrettante con 5000-6000 lire e di altrettante ancora con 6000-7000; e il rettangolo, che la rappresenta, avrà una base corrispondente alla cifra di 1834 e un'altezza tripla di quella del primo gruppo; e così via.

In questa forma di rappresentazione, il raggruppamento per categorie è causa di discontinuità e salti, che certo diminuirebbero sempre più, se la rilevazione fosse fatta per categorie comprese entro limiti ristretti e uniformi per tutte. Nel diagramma che segue, la curva punteggiata, che facciamo passare per i punti di mezzo dei lati dei rettangoli, restituisce al fenomeno la sua verosimile continuità naturale; ma è questo un procedimento empirico, che si raccomanda solo per la sua semplicità; razionalmente esso vuol essere sostituito, come si vedrà, da un processo di vera e propria interpolazione:

Si può con un facile artificio evitare la rappresentazione a rettangoli e figurare i gruppi con semplici linee, anzichè con superfici. Basta convertire, integrando, la seriazione data in quest'altra. Invece di dire: tanti individui da 1000 a 2000 lire di reddito, tant'altri da 2000 a 4000, ecc., si calcoli quanti hanno più di 1000 lire (essi nel caso in questione saranno tutti i 59.486); poi si calcoli quanti dei 59.486 posseggono più di 2000 lire (saranno evidentemente = 59.486 — 32.518 = 26.968); e ancora quanti hanno più di 4000 lire, ecc. Ciò fatto, stabilita la scala dei redditi sull'asse delle ordinate, figureremo con rette parallele all'asse delle ascisse il numero degli individui aventi un reddito superiore, rispettivamente, ai limiti di 1000 lire, di 2000, di 4000, ecc.; gli estremi di dette rette si congiungeranno con una curva. La differenza in lunghezza tra due successive parallele indicherà il numero dei componenti la classe compresa fra i corrispondenti limiti di reddito.

§ 4. Altre volte i dati forniti dall'osservazione si trovano riferiti ad una scala di misura espressa in termini qualitativi, anzichè quantitativi. La rappresentazione grafica allora non è possibile, ammenochè non si adotti un criterio accettabile per convertire la scala qualitativa in quantitativa. Abbiasi, per esempio, la classificazione per gradi degli ufficiali e sott'ufficiali della marina militare. Noi non potremmo prendere segmenti eguali sull'asse delle ordinate per rappresentare i gradi; sarebbe come ammettere *a priori* che la distanza gerarchica fra sott'ufficiale e guardiamarina è precisamente la stessa che fra tenente di vascello e capitano di corvetta o fra contr'ammiraglio e vice-ammiraglio. Come ovviare allora alla difficoltà?

Bisogna ricorrere a qualche spediente per tradurre queste distinzioni qualitative in distinzioni quantitative, ad esempio, adottando il criterio dello *stipendio* assegnato ai vari gradi, perchè appunto lo stipendio si presume in relazione stretta coll'importanza del posto gerarchico. Prenderemo adunque sull'asse delle ordinate segmenti proporzionali all'altezza della retribuzione, e il diagramma sarà facilmente costrutto. Se il criterio dello stipendio non piace, se ne potrà adottare un altro, come il numero d'anni di carriera, che occorrono in media per passare da un grado all'altro. Lo stesso artificio applicato ai gradi di altre gerarchie (amministrativa, ecclesiastica, industriale) permetterebbe la rappresentazione grafica del modo col quale gli individui si trovano ripartiti per sfere di comando e di attribuzioni in una società, e al tempo stesso darebbe un'idea approssimata dell'eguaglianza e diseguaglianza di gradi scelti da gerarchie differenti.

§ 5. Fin qui abbiamo recato illustrazioni di diagrammi a base rettilinea; ma in qualche caso si usa dare all'asse delle ascisse la forma circolare; i raggi, eventualmente prolungati oltre la circonferenza,

servono da ordinate. Questa forma può convenire nella rappresentazione di fenomeni periodici aventi un ciclo ben determinato (anno, giorno). Così il diagramma qui sopra descritto raffigura la riparti-

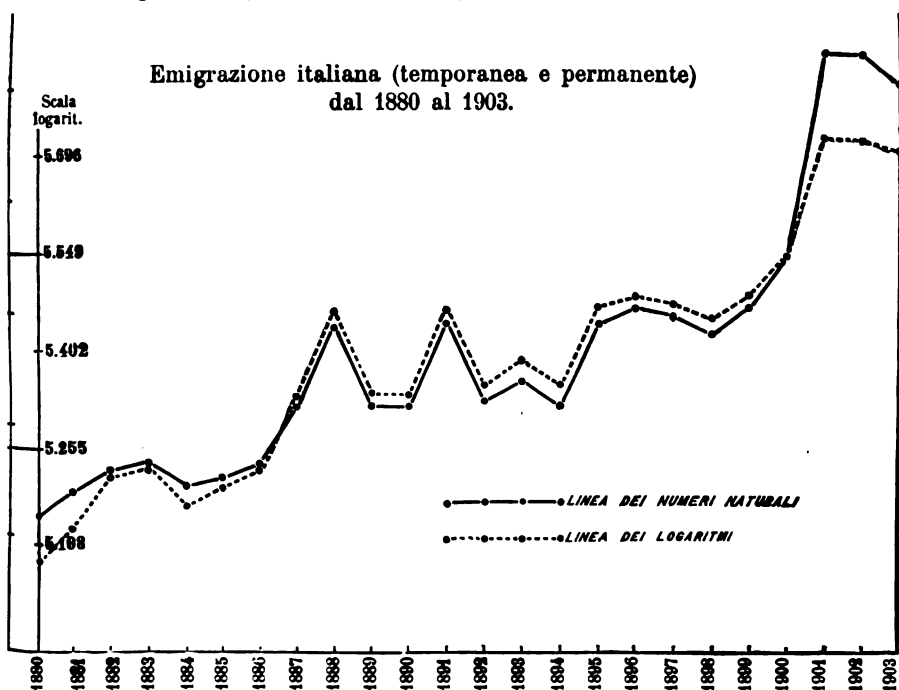
zione dei matrimoni per mesi in Italia secondo i dati del 1901. Vi si segnalano subito le eccezionali frequenze di nozze nel novembre a scapito del dicembre, nel febbraio a scapito del marzo, molti matrimoni venendo anticipati in vista dell'Avvento e della Quaresima, epoche di divieto ecclesiastico alla celebrazione solenne. Si nota pure la scarsa nuzialità dei mesi estivi.

MESI (Supposti tutti di 30 giorni)	Numero dei matrimoni supposta = 1000 la media mensile	MESI (Supposti tutti di 30 giorni)	Numero dei matrimoni supposta = 1000 la media mensile
Gennaio	1.235	Luglio	596
Febbraio. . . .	1.695	Agosto	676
Marzo	777	Settembre. . .	876
Aprile.	1.154	Ottobre. . . .	1.072
Maggio	811	Novembre. . .	1.302
Giugno	770	Dicembre . . .	1.036

Altri esempi potremmo recare di questi diagrammi a base circolare, che non sono senza qualche pratica utilità, ma che tuttavia non sostituiscono con deciso vantaggio i soliti semplici e chiari diagrammi ortogonali. Essi, fra l'altro, rendono più difficile all'occhio e alla mano il tracciare una curva perequatrice o interpolatrice.

§ 6. *Diagrammi a scala logaritmica.* — Se in luogo e vece dei dati originarii di una serie, forniti dall'osservazione, noi prendiamo i loro logaritmi adottando una conveniente scala sull'asse delle ordinate, e manteniamo per la successione dei tempi la scala naturale sull'asse delle ascisse, la costruzione del diagramma procederà nel solito modo; beninteso però che la curva descritta dal fenomeno risulterà più o meno deformata, più o meno diversa da quella che si avrebbe in una rappresentazione, tutta su scala naturale. Un andamento rettilineo, ascendente o discendente rispetto all'asse dei tempi, del fenomeno descritto col processo ordinario, si convertirebbe in un andamento curvilineo nella rappresentazione su scala logaritmica. Viceversa, se i dati della serie procedono come le successive potenze intere di un numero ($y = a^x$), la curva che ne risulterebbe su scala naturale, si convertirebbe in una retta su scala logaritmica. Fra questi due estremi possiamo immaginare infiniti casi intermedii, in cui si ottiene una modificazione più o meno grande della curva naturale.

Può convenire di procedere con scala logaritmica delle sole ordinate, quando il fenomeno è rapidamente crescente e, a parità di grandezza assoluta, le sue variazioni hanno una importanza molto minore al termine della serie che al principio. Ad es., l'emigrazione

italiana, che oscillò per qualche tempo (dal 1879 al 1886) tra la cifra di 120.000 e quella di 170.000 persone, crebbe poi con rapido passo fino a 530.000 nel 1901 e 1902. Ora è chiaro che una variazione, in più o in meno, poniamo, di 30.000 emigranti oggi non rappresenta che $^1/_{17}$ o $^1/_{18}$ della massa, laddove per lo addietro rappresentava $^1/_4$ o $^1/_5$, cioè una quota assai più rilevante. Colla figurazione grafica a scala naturale delle ordinate si ha l'inconveniente che variazioni eguali in assoluto, *sebbene diversamente importanti in via relativa*, sono tradotte nel diagramma con differenze eguali di altezza delle ordinate. Se si vuole invece che le differenze eguali di altezza delle ordinate rappresentino non variazioni assolute eguali, ma eguali variazioni *relative*, occorre adottare la scala logaritmica. Infatti se $A : B = C : D$, si avrà: $\log A - \log B = \log C - \log D$.

Per rendere comparabili in questo prospetto le due scale, naturale l'una e logaritmica l'altra, abbiamo dovuto stabilire per questa seconda due punti fissi, prendendo la media dei 12 valori più bassi della serie e quella dei 12 valori più alti. La prima media è in cifra tonda 180.000, cui corrisponde il logaritmo 5,255; la seconda media è 354.000, cui corrisponde il logaritmo 5,549. Fissati questi in coincidenza dei numeri, che loro corrispondono sulla scala naturale, non restava che prendere altri segmenti multipli o sottomultipli della differenza 5,549-5,255 e completare la scala logaritmica nel modo che si vede nel diagramma.

Ivi è facile rilevare che la curva logaritmica presenta da principio variazioni più forti della curva naturale, mentre al termine della serie ne presenta di meno forti, appunto perchè, come dicemmo, variazioni anche eguali in assoluto possono avere un'importanza relativa assai diversa, secondo che, insomma, si producono su una massa piccola di casi o su una massa già grande.

Più interessante è il caso in cui tanto i valori della funzione quanto quelli della variabile vengono sostituiti dai rispettivi logaritmi e rappresentati quindi in diagramma *a doppia scala logaritmica*. Questa forma di rappresentazione si raccomanda in modo speciale ogniqualvolta i valori della funzione diminuiscono assai rapidamente col crescere dei valori della variabile, sicchè a volerli figurare su un diagramma a scala naturale di ordinarie dimensioni, la curva da essi descritta si confonderebbe presto coll'asse delle ordinate (o delle ascisse, secondo i casi). Così il diagramma dei redditi, dato a pag. 142, presenta questo inconveniente e più lo presenterebbe se diverse fossero le curve di redditi confrontate; l'occhio non potrebbe apprezzare per i redditi più elevati il valor delle ascisse, che rappresentano il numero dei contribuenti o titolari di detti redditi. Un altro notevole vantaggio di questi diagrammi a doppia scala logaritmica sta in ciò, che alle volte riescono a sostituire una curva più semplice a quella che risulterebbe dal diagramma a scala naturale. Per esempio, una curva iperbolica dall'equazione: $y = \frac{k}{x^{\alpha}}$, su un diagramma a doppia scala logaritmica si converte in una retta; infatti:

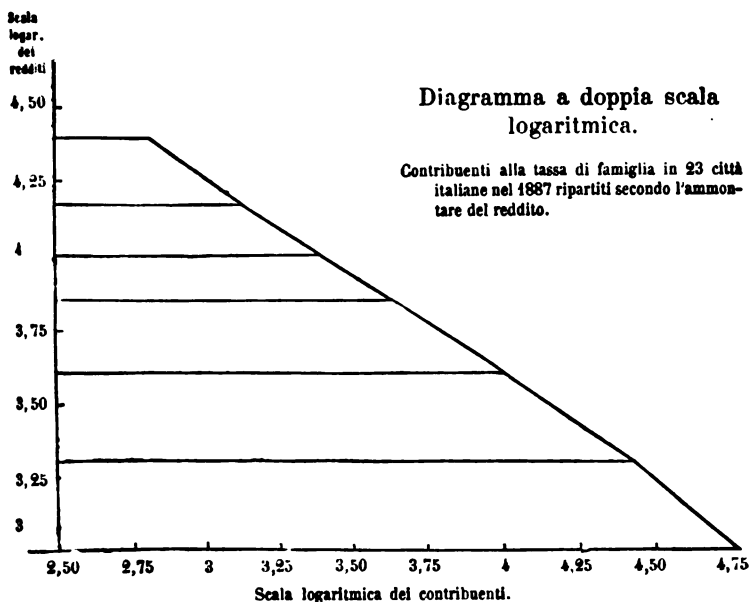
$$\log y = \log k - \alpha \log x,$$

ossia, chiamando $\log y$ con Y, $\log k$ con K, $\log x$ con X:

$$Y = K - \alpha X$$

che è precisamente l'equazione di una retta. Ora l'occhio può molto meno ingannarsi nell'apprezzare se una linea è o non è sensibilmente retta, come può invece ingannarsi nel decidere se una curva è un ramo d'iperbole, anzichè di parabola, ecc.

Proviamoci dunque a rappresentare su doppia scala logaritmica la distribuzione dei contribuenti alla tassa di famiglia, graduati per reddito, secondo i dati riferiti a pag. 141.

Sull'asse delle ordinate fisseremo la scala dei logaritmi dei redditi; cioè, invece di riportare i valori 1000 (lire), 2000, 4000, 7000, ecc., riportiamo i loro logaritmi, cioè: 3; 3,3010; 3,6021; 3,8451, ecc. Sull'asse delle ascisse, invece di riportare i numeri osservati di contribuenti che posseggono un reddito rispettivamente superiore a lire 1000, a lire 2000, ecc., riporteremo pure i logaritmi di tali numeri. Otterremo così la nuova rappresentazione, in cui la linea che congiunge le ascisse (logaritmi del numero dei contribuenti) è a un dipresso una *retta*.

DATI DELLA OSSERVAZIONE.

Redditi (x)		Contribuenti (y)
Oltre L.	1000	59.486
» »	2000	26.968
» »	4000	9.766
» »	7000	4.264
» »	10.000	2.397
» »	15.000	1.310
» »	25.000	645

LOGARITMI.

log x	log y
3	4,7744
3,3010	4,4308
3,6021	3,9897
3,8451	3,6298
4	3,3797
4,1761	3,1173
4,3979	2,8095

Per economia di spazio, si sono trascurati nel diagramma i valori di sotto di 3 nella scala delle ordinate e al disotto di 2,50 in quella lle ascisse.

La legge della seriazione viene assai meglio in evidenza in questo iagramma che in quello a scala naturale. Infatti la linea che con-

giunge le ascisse è press'a poco una retta. Quindi chiamando log y i logaritmi del numero di possessori di un reddito superiore a x, e chiamando log x i logaritmi dei redditi-limite considerati, si deve avere:

$$\log y = A - \alpha \log x.$$

Vedremo poi come si ottengono i valori di A e di α.

§ **7.** ***Rappresentazione geometrica di fenomeni a due variabili.*** — Fin qui abbiamo descritto fenomeni che sono funzione di una sola variabile; passiamo a quelli *a due variabili*. Per esempio, il numero dei nati legittimi in un dato anno per rapporto alle diverse combinazioni d'età del padre e della madre; il numero dei coscritti, secondo l'ampiezza del torace combinata colla statura; le spedizioni di merci su una rete ferroviaria, secondo la distanza e il peso, ecc.

Il Perozzo ha illustrato la ripartizione dei matrimoni in Italia, secondo le età combinate dello sposo e della sposa (1). Noi ripeteremo questo esempio, ma, per dir così, a sezione molto ridotta, l'economia dello spazio non permettendoci di riprodurre integralmente il quadro numerico, di cui si è valso il citato autore.

Indichiamo con x_1, x_2, x_3, x_4, ecc., le spose da 15 a 17 anni, da 17 a 19, da 19 a 21, da 21 a 23, ecc., e con y_1, y_2, y_3, y_4 ... gli sposi da 18 a 20 anni, da 20 a 22, da 22 a 24, da 24 a 26, ecc., seguitando così di due in due anni. E siano i matrimoni contratti così ripartiti secondo le combinazioni d'età.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	
y_1	1	3	4	3	1	1	—	—	—	—	—	13
y_2	3	8	15	13	7	3	2	1	1	—	—	54
y_3	5	16	29	31	20	10	5	2	1	1	1	121
y_4	5	18	34	40	30	17	9	5	2	1	1	163
y_5	4	15	29	36	30	20	13	6	3	2	1	159
y_6	2	9	19	24	21	17	12	7	4	2	1	118
y_7	1	5	11	15	14	12	10	6	4	2	1	81
y_8	1	3	7	10	10	9	7	5	4	3	2	61
y_9	—	2	4	6	7	6	6	4	3	3	2	43
y_{10}	—	1	2	4	4	4	4	3	3	2	2	29
y_{11}	—	—	1	2	3	3	3	3	2	2	2	21
y_{12}	—	—	—	1	2	2	2	2	2	2	1	14
	22	81	156	185	149	104	73	44	29	20	14	877

(1) V. *Annali di Statistica*, serie 3ª, vol. V.

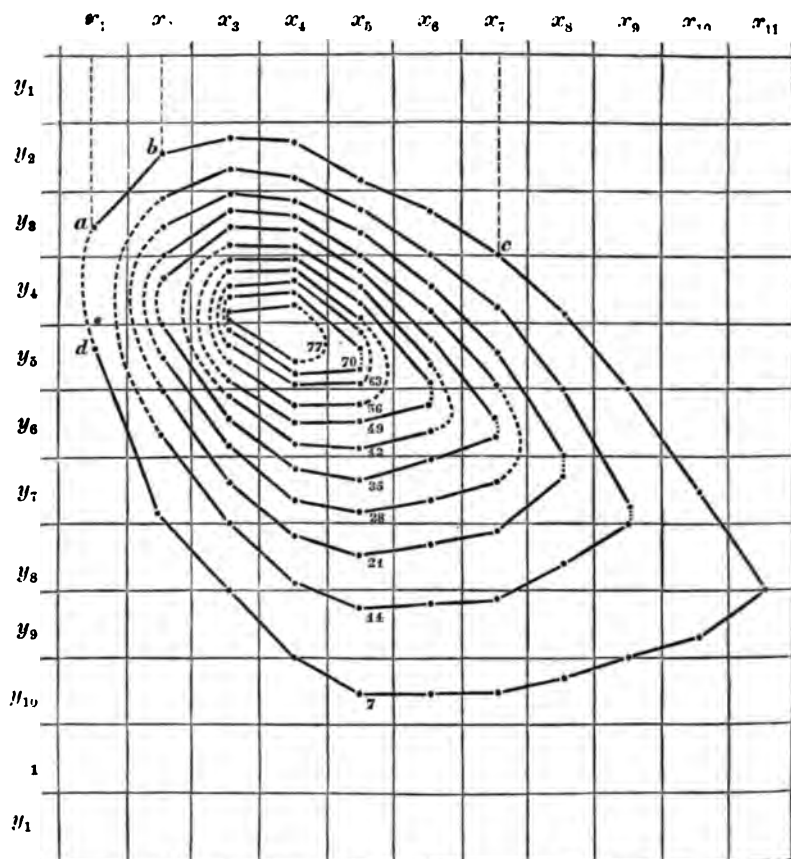
Il diagramma, col quale rappresentiamo questo prospetto numerico può ben dirsi *a curve di livello*, come se ne ha esempio nelle carte topografiche. Scelgasi per quota la cifra di 7 matrimoni. Per formare questa cifra non basta prendere le coppie di spose dell'età x_1 con sposi dell'età y_1, ma bisogna aggiungervi le coppie di spose dell'età x_1 con sposi dell'età y_2 e ancora una parte ($^3/_5$) di quelle composte di spose dell'età x_1 con sposi dell'età y_3 (1). Dunque prenderemo i primi due quadretti del diagramma e i $^3/_5$ del terzo e segneremo il punto *a*. Similmente per formare la prima quota di 7 matrimoni, in cui la donna ha l'età x_2 e l'uomo le età y_1, y_2, ecc., bisognerà prendere il primo gruppo (di 3 coppie) della 2ª colonna del prospetto numerico e aggiungervi 4 coppie delle 9, che formano il gruppo x_2 y_2; quindi segneremo il punto *b* ai $^4/_9$ del quadretto x_2 y_2, sulla linea mediana punteggiata. Così ancora per formare la prima quota di 7 matrimoni, in cui la sposa ha l'età x_7 e lo sposo ha l'età y_1, y_2, ecc., basterà prendere il primo quadretto corrispondente al gruppo x_7 y_1 (*zero* coppie), il secondo corrispondente al gruppo x_7, y_2 (= 2 coppie) e il terzo corrispondente al gruppo x_7 y_3 (= 5 coppie); sicchè il punto *c* cadrà esattamente al termine del 3° quadretto, sempre sulla linea mediana. Quanto abbiamo fatto partendo dall'alto, potremo fare partendo dal basso. I diversi punti della quota, congiunti fra loro, dànno la prima curva di livello.

La seconda curva di livello si otterrebbe in modo analogo segnando tutti i punti che corrispondono alla quota di 14 coppie (= 2 × 7), e così la terza segnando quelli che corrispondono alla quota di 21 coppie (= 3 × 7), e via di seguito.

È evidente che le successive quote a partire dall'alto o a partire dal basso non devono mai superare in valore la metà del totale della colonna numerica, su cui sono calcolate, mentre possono restare inferiori a detta metà. Così il totale della prima colonna nel prospetto numerico essendo 22, la metà sarà 11; quindi non si potrà avere nel diagramma che la quota di 7 della prima curva di livello e non già la quota 14 della seconda curva; in altre parole sulla mediana x_1 *a*, anche prolungata, non può darsi intersezione o contatto della seconda curva di livello. Ne segue però anche, nel caso concreto, che i due punti *a* e *d* si trovino distaccati, perchè la quota 7 è inferiore ad 11; la congiunzione dei due punti si può eseguire ad occhio, nel modo indicato dal diagramma.

Le aree delimitate da due successive curve di livello contengono egual numero di coppie matrimoniali.

---

(1) Diciamo per brevità $^3/_5$, sebbene non sia rigorosamente esatto. Infatti le 5 coppie del gruppo x_1 y_3 non devono essere egualmente addensate sulla superficie del 3° quadratino del diagramma; il loro addensamento va crescendo, anzichenò, quanto più si discende lungo la perpendicolare o mediana x_1 *a*. Però per una prima approssimazione basta il metodo spiccio indicato nel testo.

Nella figurazione grafica, che ha dato il Perozzo delle combinazioni d'età degli sposi in Italia, le curve ottenute somigliano ad ellissi, però non concentriche nè in tutto simili. Il nostro diagramma, sebbene calcolato su un prospetto ridotto e a cifre arrotondate, rende la stessa impressione. Le ellissi sarebbero concentriche e simili, se le combinazioni d'età avvenissero come in un giuoco di sorte, senza preferenze speciali determinate dalle somiglianze di caratteri; ma la demografia informa che tali preferenze entrano appunto in azione e modificano più o meno profondamente il risultato che si dovrebbe avere a calcolo di probabilità.

§ **8.** ***Rappresentazioni geometriche di fenomeni a tre variabili. Stereogrammi.*** — **La** posizione d'un punto, non più nel piano, ma nello ***spazio***, si determina riferendo quel punto a tre rette (***assi coordinati***) perpendicolari tra loro, passanti per una stessa *origine* e determinanti tre piani (*piani coordinati*). Ora, quando un fenomeno è funzione di tre

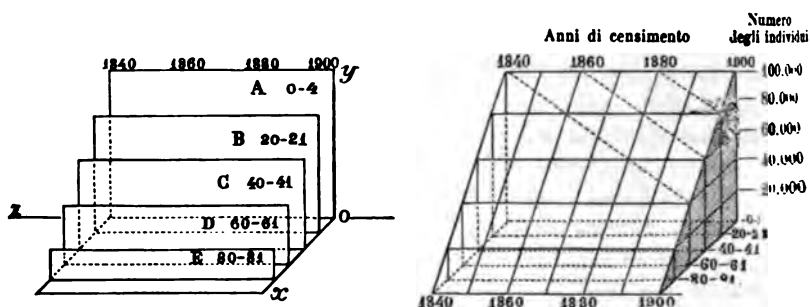
variabili (1), la rappresentazione grafica non può effettuarsene, se non mediante figure in rilievo o con modelli analoghi a quelli della geometria solida. Il nome di « stereogrammi » risponde etimologicamente giusto al concetto d'una rappresentazione in solido o a rilievo. Senonchè queste forme riescono di men facile intelligenza ed esecuzione dei diagrammi a due variabili; epperò crediamo bene di limitarci a qualche caso tipico semplicissimo, rinviando per maggiori particolari lo studioso ai saggi del Perozzo pubblicati negli *Annali di Statistica* (2).

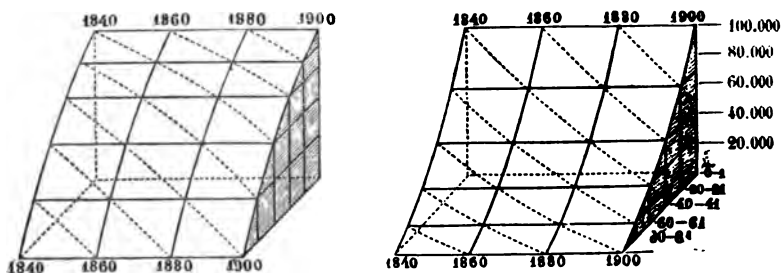
La ipotesi più semplice che si possa fare sulla composizione di una popolazione osservata in una serie di censimenti è quella di una popolazione stazionaria, i cui componenti diminuiscono di numero in progressione aritmetica col crescere dell'età anno per anno, il che equivale a dire che i morti nelle diverse età sono una cifra costante. Figuriamoci dunque disegnata su un piano la classe da 0 a 1 anno d'età; se questa classe non crebbe, nè scemò di numero in successivi censimenti (es. dal 1840 al 1900), la linea che la rappresenta sarà parallela all'asse delle ascisse; quindi avremo il rettangolo A del piccolo diagramma a sinistra, a pagina seguente. Passiamo ora alla classe da 20 a 21 anni, saltando le intermedie per non ingombrare la figura; disegnando quella sopra un secondo foglio o piano, avremo (se anche detta classe è stata stazionaria) il rettangolo B, che porremo davanti al primo a una certa distanza presa sulla ox. E così figurando su un terzo foglio o piano la classe d'età da 40 a 41, si disponga il rettangolo C verticalmente dinnanzi a B, a un intervallo eguale a quello per cui B dista da A. E così via. Se ora l'altezza dei diversi rettangoli, ossia il numero dei componenti le classi d'età rappresentate, decresce, giusta l'ipotesi fatta, in progressione aritmetica, la linea che congiunge gli estremi delle ordinate elevate sull'asse ox sarà neces-

(1) Esempi di fenomeni a tre variabili: la frequenza delle nascite legittime in rapporto all'età del padre, a quella della madre e al tempo dal quale dura il loro matrimonio; la distribuzione dei coscritti per statura, torace e indice cefalico; il movimento ferroviario di merci a piccola velocità secondo la distanza, il peso delle spedizioni e l'altezza della tariffa applicata, ecc.

(2) Veggasi L. Perozzo, *Della rappresentazione grafica di una collettività di individui nella successione del tempo e in particolare dei diagrammi a tre coordinate* (*Annali di Statistica*, serie 2ª, vol. XII, Roma, Eredi Botta, 1880). Inoltre: *Stereogrammi demografici* (*Id.*, *id.*, serie 2ª, vol. XXII, anno 1881), dove figura, tra gli altri, il noto e interessante stereogramma della popolazione svedese osservata nella lunga serie di censimenti dal 1750 al 1875.

Nella prima di queste memorie il Perozzo espone i sistemi di Berg, Knapp, Becker, Lexis e Lewin *a due assi* e quello di Zeuner *a tre assi*, ch'egli giudica preferibile ad ogni altro per la rappresentazione di una collettività di individui nella successione del tempo. Inoltre considera alcuni *tipi teorici*, rispondenti all'ipotesi di una « popolazione stazionaria » o a quella di una « popolazione progrediente in ragione aritmetica », in cui le classi di censiti o di superstiti diminuiscono di una quantità costante col crescere dell'età, anno per anno.

sariamente una retta. Si capisce quindi come dalla figura *discontinua* a sinistra (che dobbiamo immaginare completata colla rappresentazione delle classi intermedie) si passi a quella *continua* di destra. La quale ci dà l'idea di ciò che potrebbesi chiamare una « popolazione piana »; le linee dei *coetanei* vi risultano parallele e orizzontali; quelle dei *censiti*, tutte rette, eguali e parallele, tracciate dall'alto al basso

con apparente direzione da destra a sinistra (l'apparenza essendo un mero effetto di prospettiva); e quelle dei *superstiti* (di cui solo alcune per non ingenerare confusione son tracciate mediante tratteggio) ancora rette e parallele, con direzione reale da sinistra a destra, dall'alto al basso.

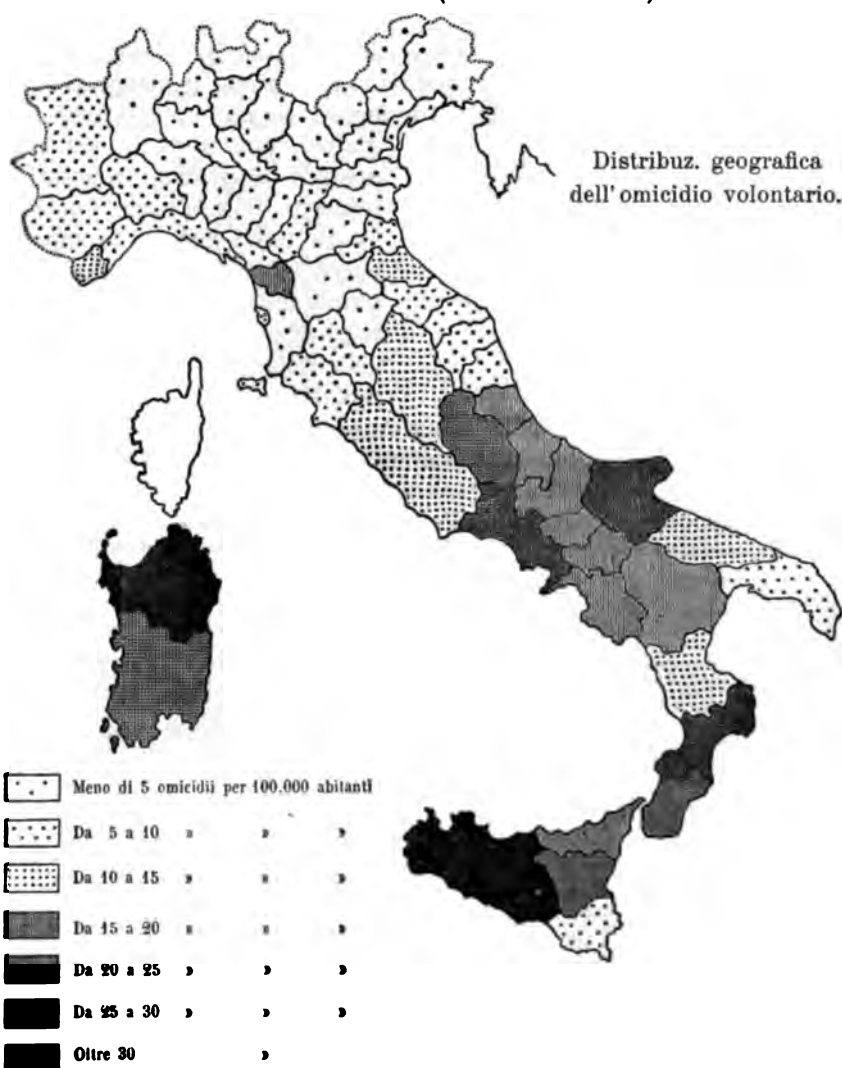
Altra ipotesi assai semplice è quella di una popolazione ancora stazionaria, ma di cui le classi diminuiscono coll'età in una progressione diversa dall'aritmetica, per esempio, seguendo una parabola di 2° grado, concava o convessa rispetto all'asse delle età. Allora le linee dei coetanei sono rette e orizzontali, quelle dei censiti e dei superstiti rami di parabole. Veggansi i due seguenti modelli:

Si può ora complicare maggiormente l'ipotesi e supporre una popolazione non più stazionaria, ma progrediente in una certa ragione aritmetica o geometrica, e di cui le classi diminuiscano secondo una data legge col crescere dell'età. La costruzione dello stereogramma sarà facilitata in ogni modo dal disegno dei diagrammi riferentisi alle singole classi (o alle principali tra esse) e dalla loro disposizione in prospettiva come nella prima figura a sinistra, non dimenticando di congiungere con una linea i punti corrispondenti dei diversi diagrammi, per dare al rilievo la necessaria continuità. Tuttavia non

crediamo di presentare qui ulteriori modelli, poichè nostro proposito era di limitarci a poche nozioni elementari in argomento, rinviando poi lo studioso ai saggi del Perozzo e alle opere ivi citate in apposita appendice.

B) *Cartogrammi.*

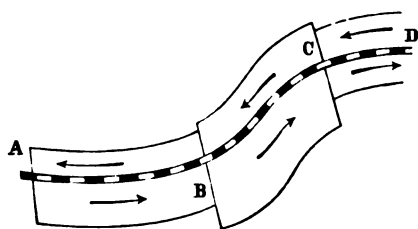
Sommario: § 1. Cartogrammi a *tinte graduate*, a *nastro*, ecc. — § 2. Vantaggi e inconvenienti delle rappresentazioni geometriche e grafiche.

§ 1. I cartogrammi sono rappresentazioni di fatti statistici su carte geografiche o su modelli, nei quali i colori colle loro gradazioni si assumono come elementi dimensivi (scale cromatiche).

Distribuz. geografica dell'omicidio volontario.

Nella precedente cartina dell'Italia è figurata la distribuzione per provincie degli omicidii volontarii d'ogni specie, secondo la media del triennio 1897-1899, per 100.000 abitanti. Le gradazioni di colore impiegate sono sette e il loro valore è indicato nella nota « Segni convenzionali ».

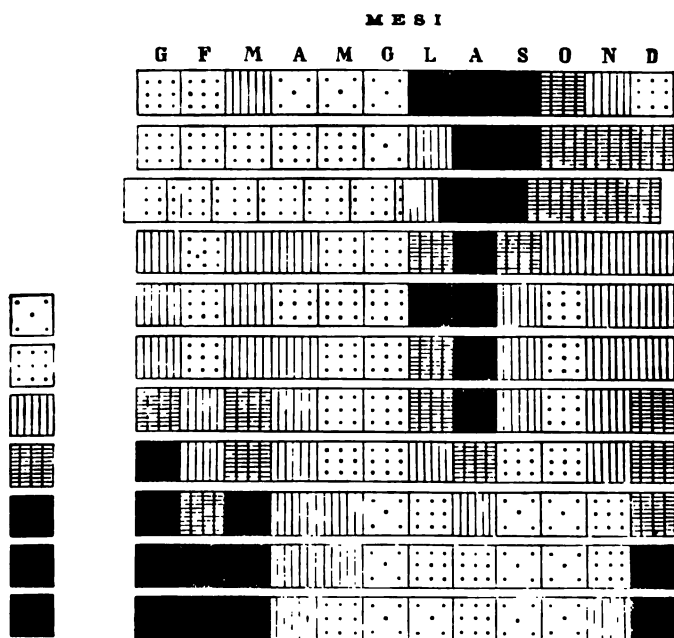
La necessità del raggruppamento in categorie, per economizzare sulle gradazioni, nasconde naturalmente diseguaglianze anche considerevoli tra provincie, alle quali fu assegnata la stessa gradazione. Così Girgenti con 40,48 omicidii per 100.000 abitanti non si distingue da Caltanissetta, che ne ha 31,73; e Treviso e Genova con le quote di 1,44 e 1,50 rispettivamente figurano alla pari con Firenze, che ha la quota di 4,82.

I cartogrammi di queste specie si dicono *a tinte graduate*.

I cosidetti cartogrammi *a nastro*, che si usano per rappresentare sulle carte geografiche il movimento ferroviario delle merci e dei viaggiatori, non sono in realtà che speciali diagrammi adattati al tracciato delle ferrovie. Così nella figura qui a fianco, le superfici curve adiacenti alla linea ferroviaria indicano, secondo la loro larghezza e secondo la direzione delle freccie, il traffico dei tronchi AB, BC, CD.

Diamo ora un esempio di cartogramma non geografico. Sia da rappresentare la ripartizione per mesi dei morti delle varie età, in Italia, secondo i dati del 1886, fatta = 100 la media mensile per ogni classe d'età e supposti i mesi formati d'egual numero di giorni:

ETÀ	Gennaio	Febbraio	Marzo	Aprile	Maggio	Giugno	Luglio	Agosto	Settembre	Ottobre	Novembre	Dicembre
1-5	84	93	94	80	76	81	117	145	135	115	96	87
5-10	90	87	92	91	84	82	98	132	122	108	107	107
10-20	87	85	92	88	92	85	111	131	116	105	104	103
20-30	95	89	96	94	91	90	112	123	111	98	101	101
30-40	94	88	100	92	90	91	118	134	105	93	96	98
40-50	101	92	99	96	89	90	116	126	103	91	95	101
50-60	110	101	106	95	92	85	107	117	94	87	96	109
60-70	120	105	114	96	93	83	101	107	88	85	95	112
70-80	132	114	125	98	94	81	91	96	82	79	92	115
80-90	143	118	124	95	96	78	87	91	80	78	92	118
Oltre i 90 . .	151	120	125	104	86	74	81	88	71	79	103	119

La semplice ispezione delle cifre mostra il prevalere della mortalità dei bambini nei mesi estivi, dei vecchi nei mesi invernali. Si noti peraltro che abbiamo lasciati fuori conto i bambini dalla nascita a un anno, tra i quali la mortalità è notevole pur nell'inverno. Per tutte le età i mesi più propizii sono generalmente l'aprile, il maggio, il giugno e l'ottobre. Ora, per l'esecuzione di un cartogramma (volendo noi valerci soltanto del nero e di alcune gradazioni di grigio) procederemo nel seguente modo: Si figurino le singole classi d'età rappresentate da listerelle rettangolari divise in 12 parti, quanti cioè i mesi dell'anno. Ognuna di queste parti porterà la gradazione di colore che le spetta in ragione dell'intensità relativa della mortalità. All'uopo, facciamo la differenza tra il numero più alto dato dalla precedente tabella (151) e il più basso (71); tale differenza è di 80 punti. Se 7 sono le gradazioni di colore disponibili, dividendo 80 per 7, si ottiene il quoziente 11,43. Allora a partire da 71 avremo:

Da	71 - 82,4	. . . . .	I gradazione
»	82,4- 93,8	. . . . .	II »
»	93,8-105,3	. . . . .	III »
»	105,3-116,7	. . . . .	IV »
»	116,7-128,1	. . . . .	V »
»	128,1-139,5	. . . . .	VI »
»	139,5-151	. . . . .	VII »

Mortalità relativa per mesi e per classi d'età in Italia nel 1886.

Molt'altre forme di cartogrammi sono in uso, ma non hanno qui bisogno di una speciale illustrazione, essendo tutte di facilissima intelligenza.

§ 2. Quel che abbiamo detto fin qui mette in chiaro i vantaggi dei procedimenti geometrici e grafici. Essi rendono agevole all'occhio di cogliere a volo l'andamento generale di un fenomeno, i suoi caratteri di periodicità, continuità o discontinuità, ecc., il suo comportarsi secondo certe forme matematiche semplici, di cui l'interpolazione fornisce poi l'espressione analitica; pongono in risalto le variazioni concomitanti dei fenomeni paragonati e riescono suggestivi nell'indagine delle cause. Non vanno peraltro esenti da qualche inconveniente. Così le variazioni di piccola ampiezza, che pur possono avere la loro importanza, tendono a scomparire nei diagrammi costruiti su scale non grandi, venendo in certo modo sacrificate alla comodità del formato; scompaiono anche più facilmente nei cartogrammi a tinte graduate, quando per economia si debba dare scarso sviluppo alla scala cromatica. Fu avvertito pure un aspetto arbitrario, che diminuisce i pregi delle rappresentazioni geometriche. Se chi fa la rappresentazione ha interesse a mettere in evidenza certe variazioni, costruirà un diagramma ad angoli molto acuti, prendendo una piccola scala di misura per i valori della variabile e una grande per i valori della funzione; se ha l'interesse opposto, costruirà un diagramma ad angoli ottusi, prendendo una piccola scala pei valori della funzione e una grande per quelli della variabile. Ma anche a questo arbitrio si può rimediare per convenzione, collo stabilire le scale delle ordinate e delle ascisse in modo che una variazione di media grandezza del fenomeno dia luogo, per esempio, ad un angolo di 45°.

Ad ogni modo l'enorme materiale statistico, che le amministrazioni pubbliche ed i privati dei paesi civili vanno continuamente ammassando, non potrà in un prossimo avvenire prestarsi bene alle indagini degli studiosi se non nelle forme geometriche e grafiche di rappresentazione e meglio ancora nelle forme analitiche, che sono l'argomento del titolo seguente.

TITOLO III.

PROCEDIMENTI ALGEBRICI.

A) *Interpolazione di serie e seriazioni per via di funzioni algebriche intere.*

Sommario: § 1. Diversi uffici dell'interpolazione. — § 2. Interpolazioni lineari e paraboliche. — § 3. Scomposizione di elementi raggruppati in classi. — § 4. Valore della variabile, che bipartisce esattamente una classe. — § 5. Interpolazione col metodo dei *minimi* quadrati. — § 6. Esempio pratico e tabella di valori.

§ 1. Dai procedimenti aritmetici, geometrici e grafici, coi quali si mira a rappresentare e semplificare le serie e seriazioni, passiamo ai *procedimenti algebrici*, in particolare al problema dell'*interpolazione*.

L'« interpolazione » è una operazione alla quale si ricorre nei seguenti casi:

1° quando si vogliano inserire dati a calcolo fra i dati rilevati a intervalli più o men grandi per un fenomeno di sua natura continuativo. Ad esempio, sapendosi quant'era la popolazione italiana nel 1871, nel 1881 e nel 1901, calcolare quanta dovette essere negli anni intermedii a questi, nell'ipotesi che essa abbia seguito un andamento parabolico;

2° allorchè si tratti di scomporre nei loro elementi i dati di una serie, che la rilevazione ci fornisce raggruppati in classi più o meno estese. Così, sapendosi che in un dato paese i censiti fra 20 e 25 anni d'età furono 200.000, quelli da 25 a 30, 190.000, e quelli in età da 30 a 35, 177.000, calcolare quanti probabilmente erano in età da 20 a 21, da 21 a 22, da 22 a 23, ecc.;

3° quando, per comodità di studio, convenga sostituire alla curva complicata fornita dall'osservazione una curva più semplice, che tenga conto dell'andamento generale della serie e delle sue principali ondulazioni, prescindendo dalle più minute oscillazioni, che si presumono dovute a cause minute, accidentali, inaccessibili all'analisi. In questo terzo caso, che è di gran lunga il più importante, noi miriamo a ricavare dalla curva empirica la vera legge del fenomeno o almeno una espressione approssimata, per quanto è possibile, di quella che dovrebbe essere la legge naturale del fenomeno. E se nel capitolo relativo ai procedimenti geometrici abbiamo trattata la questione di rappresentare con una curva una data funzione algebrica o trigonometrica, qui invece dovremo occuparci della questione inversa, cioè, data una curva empirica, determinarne l'equazione approssimata quanto si desidera.

§ **2.** Supponiamo di dover tra *due* valori dati interpolare quanti termini si vogliano, a intervalli eguali, in modo che alla rappresentazione grafica essi costituiscano punti situati su una medesima *retta*. Oppure fra *tre* valori dati inserire quanti termini si vogliano, ad intervalli eguali, in modo che le ordinate, che li rappresentano, abbiano gli estremi situati su una *parabola di secondo grado*. O ancora: fra *quattro* valori dati, interpolare quanti termini si vogliano, in serie rappresentabile graficamente con una curva di 3° grado, ecc.

La soluzione di questi quesiti crediam bene di facilitarla con pratici esempi.

Sia la popolazione di un paese aumentata da 100 a 115 nel corso di un quinquennio. Supponendo che abbia seguito un movimento rettilineo, per sapere quanto essa dovette essere alla fine del primo, del secondo, del terzo e del quarto anno del quinquennio, disporremo la serie nella maniera indicata dalla funzione:

$$y = ax + b$$

che noi già sappiamo essere l'equazione di una retta, dando ad x i successivi valori 0, 1, 2, 3, 4, 5. Con che si ottiene:

$$\begin{array}{lll}
\text{Per } x = 0, & y = \phantom{2a +{}} b = 100 & \\
\text{»} \; x = 1, & y = a + b & \\
\text{»} \; x = 2, & y = 2a + b & \text{Termini} \\
\text{»} \; x = 3, & y = 3a + b & \text{da inserire} \\
\text{»} \; x = 4, & y = 4a + b & \\
\text{»} \; x = 5, & y = 5a + b = 115 &
\end{array}$$

Ora, essendo $b = 100$; $5a + b = 115$, si determina facilmente il valore di $a = 3$. Quindi il primo termine da inserire, ossia $a + b$, sarà eguale a 103; il secondo, ossia $2a + b = 106$; il terzo, $3a + b = 109$; il quarto, $4a + b = 112$. I numeri così determinati sono in progressione aritmetica, ossia le loro differenze prime sono una costante (= 3); graficamente darebbero luogo a punti situati su una medesima retta. L'equazione di questa retta è dunque:

$$y = 3x + 100$$

l'origine delle coordinate essendo fissata al principio del quinquennio. Era del resto in nostro arbitrio fissare l'origine delle coordinate in uno qualunque degli anni del quinquennio, anzichè al principio. Per esempio, se l'avessimo fissata alla fine del secondo anno, i valori successivi attribuiti alla variabile x sarebbero stati i seguenti: — 2, — 1, 0, + 1, + 2, + 3; e la serie sarebbe riuscita disposta così:

$$\begin{array}{lll}
\text{Per } x = -2, & y = -2a + b = 100 & \\
\text{»} \; x = -1, & y = -a + b & \\
\text{»} \; x = 0, & y = \phantom{2a +{}} b & \text{Termini} \\
\text{»} \; x = 1, & y = a + b & \text{da inserire} \\
\text{»} \; x = 2, & y = 2a + b & \\
\text{»} \; x = 3, & y = 3a + b = 115 &
\end{array}$$

Allora essendo: $-2a + b = 100$; $3a + b = 115$, si ricava facilmente: $a = 3$, $b = 106$.

L'equazione della retta, *colla origine alla fine del secondo anno*, diventa:

$$y = 3x + 106.$$

Sarebbe con ciò mutata la costante b; ma i valori interpolati risultano ancora come poc'anzi, semprechè si facciano assumere alla variabile le grandezze da -2 a $+3$; vale a dire si ottiene ancora la serie: 103, 106, 109, 112 per i quattro termini che ci proponevamo di inserire.

Facciamo ora il caso di una interpolazione parabolica di 2° grado. Nel 1890, per es., la popolazione di un dato paese era 100, nel 1895 si trovò essere 115, nel 1900 salì a 122. Si domanda: quanta dovette essere negli anni intermedii a quelli indicati, nell'ipotesi che il suo movimento abbia seguito una curva parabolica ordinaria?

Basterà porre la serie sotto la forma

$$y = ax^2 + bx + c$$

facendo assumere alla variabile x i valori successivi -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5 per comodità di calcolo (1). Le espressioni della forma $y = ax^2 + bx + c$ sono appunto le equazioni di parabole ordinarie o di 2° grado. Si avrà dunque:

Anni	Popolazione	
1890	$25a - 5b + c = 100$	
1891	$16a - 4b + c$	Termini da inserire
1892	$9a - 3b + c$	
1893	$4a - 2b + c$	
1894	$a - b + c$	
1895	$c = 115$	
1896	$a + b + c$	Termini da inserire
1897	$4a + 2b + c$	
1898	$9a + 3b + c$	
1899	$16a + 4b + c$	
1900	$25a + 5b + c = 122$	

Abbiamo insomma questo semplice sistema di equazioni:

$$25a - 5b + c = 100$$
$$c = 115$$
$$25a + 5b + c = 122$$

donde si ricavano i valori:

$$a = -0,16; \quad b = 2,20; \quad c = 115.$$

---

(1) Infatti, ponendo l'origine delle coordinate alla metà esatta della serie, si ottengono gli stessi coefficienti di a e di b per punti situati ad egual distanza dall'origine, il che semplifica di molto la soluzione del sistema di equazioni.

E applicando tali valori, si ha la serie seguente, di cui calcoliamo anche le differenze prime e seconde:

Anni	Popolazione	Differenze prime	Differenze seconde
1890 . . .	$25a - 5b + c = 100$		
1891 . . .	$16a - 4b + c = 103,64$	3,64	— 0,32
1892 . . .	$9a - 3b + c = 106,96$	3,32	— 0,32
1893 . . .	$4a - 2b + c = 109,96$	3,00	— 0,32
1894 . . .	$a - b + c = 112,64$	2,68	— 0,32
1895 . . .	$c = 115$	2,36	— 0,32
1896 . . .	$a + b + c = 117,04$	2,04	— 0,32
1897 . . .	$4a + 2b + c = 118,76$	1,72	— 0,32
1898 . . .	$9a + 3b + c = 120,16$	1,40	— 0,32
1899 . . .	$16a + 4b + c = 121,24$	1,08	— 0,32
1900 . . .	$25a + 5b + c = 122$	0,76	

Caratteristica di queste serie paraboliche di 2° grado è che le differenze prime tra i successivi termini costituiscono una progressione aritmetica (cioè una serie *lineare*) e le differenze seconde (differenze delle differenze) sono una quantità costante.

L'equazione della curva descritta coi tre numeri dati e con quelli interpolati è dunque:

$$y = -0,16\, x^2 + 2,20\, x + 115$$

riferita l'origine all'anno 1895 (1).

Anche in questo caso era in nostro arbitrio la scelta dell'anno da fissare come origine delle coordinate. Potevamo, per esempio, stabilire il 1892 invece del 1895. Allora la variabile x, invece di assumere i successivi valori da (— 5) a (+ 5), avrebbe assunto quelli da (— 2) a (+ 8); e l'equazione della parabola sarebbe risultata:

$$y = -0,16\, x^2 + 3,16\, x + 106,96$$

dalla quale si ricavano, tenuto conto della nuova origine assegnata alle coordinate, i medesimi valori che dalla precedente.

Analogamente procederemmo se, dati quattro termini dell'osservazione, volessimo inserire altri termini costituenti con quelli una serie rappresentabile con una curva di 3° grado, cioè una serie i cui termini

---

(1) L'espressione $ax^2 + bx + c$ può facilmente trasformarsi in quest'altra: $a(x + h)^2 + k$, dove $h = \frac{b}{2a}$ e $k = c - \frac{b^2}{4a}$; la quale trasformazione in casi particolari offre il vantaggio di una maggiore semplicità. Sicchè invece di:

$$y = -0,16x^2 + 2,20x + 115$$

possiamo scrivere:

$$y = 0,16\,(x + 6,875)^2 + 122,5625.$$

abbiano per differenze terze una quantità costante. Bisognerebbe porre la serie sotto la forma $y = ax^3 + bx^2 + cx + d$, dando al solito alla variabile successivi valori nell'ordine dei numeri naturali, anche a cominciare da quantità negative e preferibilmente fissando l'origine delle coordinate a metà precisa della serie.

Siano, ad es., i quattro numeri 50, 56, 60, 70, tra cui si vogliano inserire tre termini costituenti con essi una serie rappresentabile con una curva di 3° grado. Stabiliremo il sistema di equazioni nel modo indicato qui sotto, utilizzando le quattro equazioni complete:

$$\begin{aligned} -27a + 9b - 3c + d &= 50 \\ -8a + 4b - 2c + d &= \dots \\ -a + b - c + d &= 56 \\ d &= \dots \\ a + b + c + d &= 60 \\ 8a + 4b + 2c + d &= \dots \\ 27a + 9b + 3c + d &= 70 \end{aligned}$$

da cui si ricavano i seguenti valori:

$$a = 0{,}1666\dots; \quad b = 0{,}25; \quad c = 1{,}8333\dots; \quad d = 57{,}75.$$

I numeri interpolati, in grassetto, completano la serie che si riconosce parabolica di 3° grado perchè le differenze terze dei termini sono una quantità costante.

	Differenze prime	Differenze seconde	Differenze terze
$-27a + 9b - 3c + d = 50$			
	3,75		
$-8a + 4b - 2c + d =$ **53,75**		− 1,50	
	2,25		+ 1
$-a + b - c + d = 56$		− 0,50	
	1,75		+ 1
$d =$ **57,75**		+ 0,50	
	2,25		+ 1
$a + b + c + d = 60$		+ 1,50	
	3,75		+ 1
$8a + 4b + 2c + d =$ **63,75**		+ 2,50	
	6,25		
$27a + 9b + 3c + d = 70$			

L'equazione della curva rappresentatrice sarebbe:

$$0{,}1666\, x^3 + 0{,}25\, x^2 + 1{,}8333\, x + 57{,}75 \text{ (1).}$$

§ 3. Altro ufficio dell'interpolazione dicemmo essere quello di scomporre nei loro elementi i dati di una serie, che la rilevazione ci fornisce raggruppati in classi più o meno estese.

Siano i censiti in un paese nell'età da 40 a 45 anni in numero di 300 mila, quelli in età da 40 a 50, 280 mila, e quelli da 50 a 55 anni,

---

(1) Anche l'espressione $ax^3 + bx^2 + cx + d$ è trasformabile in quest'altra: $a(x + h)^3 + k$, dove h e k sono funzioni speciali di a, b, c e d; ma la trasformazione riesce piuttosto laboriosa.

250 mila. Si domanda: quanti saranno probabilmente gli individui in età da 40 a 41, da 41 a 42, da 42 a 43, ecc.?

Sarebbe uno spediente grossolano quello di dividere per 5 ogni gruppo quinquennale e attribuire a ciascuna classe annuale risultante un numero eguale di censiti; si avrebbe, tra l'altre cose, l'inconveniente di un brusco salto nel passaggio dall'ultimo anno di una data classe quinquennale al primo della classe successiva, laddove è da supporre piuttosto un decrescere continuo e regolare della popolazione in funzione dell'età. Perciò, attenendoci a questa ipotesi, è preferibile procedere per interpolazione parabolica, ponendo la serie dei censiti, anno per anno d'età, sotto la solita forma:

$$y = ax^2 + bx + c$$

coll'origine fissata, per comodità di calcolo, all'età da 47 a 48 anni, che è centrale per l'insieme dei tre gruppi. Avendosi dunque 15 classi annuali raccolte in tre quinquennali, la variabile x assumerà i valori da (— 7) a (+ 7). Si avrà cioè:

Età	N° dei censiti	Somme
40-41 . . .	$49a - 7b + c$	
41-42 . . .	$36a - 6b + c$	
42-43 . . .	$25a - 5b + c$	$135a - 25b + 5c = 300.000$
43-44 . . .	$16a - 4b + c$	
44-45 . . .	$9a - 3b + c$	
45-46 . . .	$4a - 2b + c$	
46-47 . . .	$a - b + c$	
47-48 . . .	c	$10a + 5c = 280.000$
48-49 . . .	$a + b + c$	
49-50 . . .	$4a + 2b + c$	
50-51 . . .	$9a + 3b + c$	
51-52 . . .	$16a + 4b + c$	
52-53 . . .	$25a + 5b + c$	$135a + 25b + 5c = 250.000$
53-54 . . .	$36a + 6b + c$	
54-55 . . .	$49a + 7b + c$	

Si ottiene così un sistema di tre equazioni con tre incognite:

$$135a - 25b + 5c = 300.000$$
$$10a + 5c = 280.000$$
$$135a + 25b + 5c = 250.000$$

la cui facile risoluzione fornisce i seguenti valori:

$$a = -40; \quad b = -1000; \quad c = 56.080$$

Applicando detti valori nelle formole che esprimono le singole classi annuali d'età, si ottiene la serie parabolica che diamo a pagina seguente, nella qual serie è facile verificare che le differenze prime tra i successivi termini costituiscono una progressione aritmetica, mentre le differenze seconde sono costanti ed eguali ad 80.

ETÀ	N° dei censiti	ETÀ	N° dei censiti	ETÀ	N° dei censiti
40-41 . . .	61.120	45-46 . . .	57.920	50-51 . . .	52.720
41-42 . . .	60.640	46-47 . . .	57.040	51-52 . . .	51.440
42-43 . . .	60.080	47-48 . . .	56.080	52-53 . . .	50.080
43-44 . . .	59.440	48-49 . . .	55.040	53-54 . . .	48.640
44-45 . . .	58.720	49-50 . . .	53.920	54-55 . . .	47.120
Somma	300.000	*Somma*	280.000	*Somma*	250.000

§ 4. Qui si presenta l'opportunità di trattare il quesito: *data una serie di elementi raggruppati in classi più o meno comprensive, determinare quel valore della variabile che bipartisce esattamente una qualunque di tali classi.*

Senza ricorrere a nozioni di calcolo infinitesimale, ci contenteremo della soluzione algebrica *per successive approssimazioni.*

Riprendiamo l'esempio precedente e domandiamoci qual'è l'età, in anni e frazioni d'anno, che divide in due parti eguali il gruppo di 280.000 individui censiti in età da 45 a 50, fermo restando il carattere parabolico (di 2° grado) della seriazione. Ora, per formare la cifra di 140.000, cioè la metà precisa del gruppo in parola, non bastano le classi da 45 a 46 e da 46 a 47 anni, dianzi decomposte coll'interpolazione; chè esse darebbero un complesso di 114.960 individui appena; d'altra parte, ce ne sarebbe d'avanzo, se vi aggregassimo la intera classe da 47 a 48. Evidentemente il valore della variabile, che bipartisce il gruppo quinquennale, deve trovarsi entro i limiti da 47 a 48. Tenteremo allora una prima approssimazione, suddividendo la classe da 47 a 48 pure in cinque parti (o anche più, se così piace), cioè frazionandola nelle seguenti sottoclassi, che procedono di due decimi in due decimi di anno: 47,0-47,2; 47,2-47,4; 47,4-47,6; 47,6-47,8; 47,8-48,0. S'immagini dunque disposto lo sviluppo della funzione come nel paragrafo precedente:

Classi annuali d'età		Formole		Numero degli individui
46-47 (1)			$135a' - 25b' + 5c'$	$= 57.040$
47-48	47,0-47,2	$4a' - 2b' + c'$		
	47,2-47,4	$a' - b' + c'$		
	47,4-47,6	c'	$= 10a' + 5c'$	$= 56.080$
	47,6-47,8	$a' + b' + c'$		
	47,8-48,0	$4a' + 2b' + c'$		
48-49 (1)			$135a' + 25b' + 5c'$	$= 55.040$

(1) Di queste due classi annuali non indichiamo, per brevità, la suddivisione in cinque sottoclassi.

Sistema d'equazioni, da cui si ricavano i valori:

$$a' = -0{,}32;\quad b' = -40;\quad c' = 11.216{,}64.$$

Quindi, applicandoli alle formole concernenti la suddivisione della classe 47-48, avremo:

Suddivisioni della classe 47-48		Numero degli individui	
47,0-47,2	$= 4a' - 2b' + c'$	$= 11.295{,}36$	
47,2-47,4	$= a' - b' + c'$	$= 11.256{,}32$	
47,4-47,6	$= c'$	$= 11.216{,}64$	$= 56.080$
47,6-47,8	$= a' + b' + c'$	$= 11.176{,}32$	
47,8-48,0	$= 4a' + 2b' + c'$	$= 11.135{,}36$	

Per avvicinarci dunque alla cifra di 140.000, esatta metà del gruppo quinquennale 45-50, prenderemo le seguenti classi e sottoclassi:

45-46	= 57.920
46-47	= 57.040
47,0-47,2	= 11.295,36
47,2-47,4	= 11.256,32
Totale	137.511,68

Il risultato è in difetto in confronto della cifra che si vuol raggiungere; ma se avessimo messo in conto anche la sottoclasse 47,4-47,6, quella cifra sarebbe stata oltrepassata. Il valore della variabile, che bipartisce esattamente il gruppo 45-50, è dunque compreso fra i limiti 47,4 e 47,6.

Allora si tenti una seconda approssimazione, suddividendo ancora la sottoclasse 47,4-47,6 in altre cinque parti, che evidentemente procederanno di quattro in quattro centesimi di anno:

Sottoclassi				
47,2-47,4	. . . .		$= 135a'' - 25b'' + 5c''$	$= 11.256{,}32$
47,4-47,6	47,40-47,44	$4a'' - 2b'' + c''$		
	47,44-47,48	$a'' - b'' + c''$		
	47,48-47,52	c''	$= 10a'' + 5c''$	$= 11.216{,}64$
	47,52-47,56	$a'' + b'' + c''$		
	47,56-47,60	$4a'' + 2b'' + c''$		
47,6-47,8	. . . .		$= 135a'' + 25b'' + 5c''$	$= 11.176{,}32$

donde si ricava: $a'' = -0{,}00256$; $b' = -1{,}6$; $c'' = 2.243{,}33$ (1).

---

(1) Il metodo, in apparenza lungo, si abbrevia, passando senz'altro dalle costanti delle equazioni primitive a quelle successive con l'uso di queste formole:

$$a' = 0{,}008\,a;\quad a'' = 0{,}008\,a';\ \text{ecc.};\qquad b' = 0{,}04\,b;\quad b'' = 0{,}04\,b';\ \text{ecc.}$$

$$c' = 0{,}2\,c - 0{,}016\,a;\quad c'' = 0{,}2\,c' - 0{,}016\,a';\ \text{ecc.}$$

Applicando i quali valori otteniamo:

Suddivisioni della sottoclasse 47,4-47,6		Numero degli individui	
47,40-47,44	$= 4a'' - 2b'' + c''$	$= 2.246,52$	
47,44-47,48	$= a'' - b'' + c''$	$= 2.244,93$	
47,48-47,52	$= c''$	$= 2.243,33$	$= 11.216,63$
47,52-47,56	$= a'' + b'' + c''$	$= 2.241,73$	
47,56-47,60	$= 4a'' + 2b'' + c''$	$= 2.240,12$	

Sicchè ci avvicineremo ancor più alla cifra di 140.000, esatta metà del gruppo quinquennale 45-50, addizionando le seguenti classi e sottoclassi:

45-46	= 57.920
46-47	= 57.040
47,0 -47,2	= 11.295,36
47,2 -47,4	= 11.256,32
47,40-47,44	= 2.246,52
Totale	139.758,20

Il risultato è in difetto di sole 242 unità in confronto della cifra da raggiungere; mentre, se avessimo messo in conto la suddivisione 47,44-47,48, sarebbe riuscito in eccesso di 2.003.

Quindi il valore della variabile, che bipartisce esattamente il gruppo quinquennale 45-50, è certamente compreso fra 47,44 e 47,48, e inoltre deve trovarsi assai più vicino al primo limite che al secondo.

Si potrebbe così proseguire con una terza o una quarta approssimazione; ma già al punto, cui siamo arrivati, la porzione d'area rappresentata dalla suddivisione 47,44-47,48 si confonde quasi con un rettangolo, la cui larghezza, misurata sull'asse delle età è di $^4/_{100}$ di anno e il cui contenuto è di 2.245 unità circa. Di queste, per completare la cifra di 140.000, ce ne abbisognano 242. Istituiremo dunque la proporzione:

$$2.245 : {}^4/_{100} = 242 : x$$
$$x = 0,43\ \%.$$

E il valore cercato sarà approssimativamente 47,4443.

§ 5. Passiamo al terzo ufficio dell'interpolazione, quello di sostituire alla curva empirica, di solito complicata da oscillazioni secondarie imputabili a cause accidentali, una curva più semplice, che si presume rappresentare la vera legge del fenomeno. Il problema consiste allora nel far passare una retta o una curva di un certo ordine, non *per* certi punti, ma *fra* certi punti, a condizioni determinate.

È evidente che, dati *tre* o *più* numeri di una serie, se le loro differenze prime non sono una quantità costante (cioè se essi già non costituiscono una progressione aritmetica), gli estremi delle ordinate che li rappresentano su un diagramma cartesiano non possono trovarsi situati su una medesima retta. Quindi è vano cercare di far passare una retta *per*

questi punti, mentre si possono tracciare infinite rette *fra* questi stessi punti. In tal caso però l'arbitrio della scelta non avrebbe limiti; bisogna che ci imponiamo noi stessi una condizione che disciplini l'arbitrio. Si conviene generalmente di assumere per condizione la seguente: che la *somma dei quadrati degli scostamenti fra i numeri teorici da trovare e i numeri forniti dall'osservazione sia la minima possibile*. Ciò non vieta che si possano adottare altre condizioni a piacere; per esempio, quella di rendere minima la somma dei cubi (supposti tutti di segno positivo) o delle quarte potenze, ecc.; ma allora i calcoli riuscirebbero più laboriosi.

È pure noto che, dati *quattro* o *più* numeri di una serie, se le loro differenze seconde non sono una quantità costante, gli estremi delle ordinate, che li rappresentano, non possono essere situati su una parabola di secondo grado. Quindi è vano cercare di far passare una curva di tale specie *per* questi punti; mentre si possono tracciare quante si vogliono parabole di secondo grado *fra* questi medesimi punti. Senonchè allora l'arbitrio della scelta non avrebbe limiti e noi dobbiamo disciplinarlo con una condizione identica a quella sopra ricordata: che la somma dei quadrati degli scostamenti fra i numeri della serie parabolica da trovare e quelli forniti dall'osservazione sia un minimo in confronto di quella che si avrebbe per qualsiasi altra serie parabolica.

Questa stessa condizione ci imponiamo, quando si tratti di far passare una parabola di terzo grado *fra cinque* o *più* punti dati. E così via, nel caso di interpolazioni paraboliche di grado più elevato.

Ci limiteremo qui ad indicare un metodo pratico per eseguire coteste interpolazioni, fornendo un prontuario di coefficienti e un esempio concreto di applicazione, senza tuttavia entrare nella dimostrazione matematica, che presuppone conoscenza del calcolo infinitesimale. Basterà che lo studioso comprenda bene come, data una serie statistica, sia sempre possibile determinare una curva teorica, la quale si avvicini fin che si vuole alla curva dell'osservazione, mediante un processo di successive approssimazioni. Così, nel caso di serie non periodiche, si comincia dall' interpolare una *retta parallela all'asse delle ascisse*, cioè dal sostituire ai dati della serie la *media aritmetica* loro; indi s' interpola una retta più o meno inclinata rispetto all'asse delle ascisse, cioè si sostituisce alla serie empirica una progressione aritmetica, con che si fa un guadagno nell'approssimazione; poi s'interpola una parabola ordinaria, ai dati dell'osservazione sostituendosi certi numeri teorici, le cui differenze seconde sono costanti, e si consegue un ulteriore guadagno di approssimazione; e così via.

§ 8. *Esempio pratico di interpolazione di serie non periodiche col metodo dei minimi quadrati.* — La formola generale, di assai facile uso quando la serie osservata procede per intervalli eguali di tempo, è:

$$y = A + B\psi_{I} + C\psi_{II} + D\psi_{III} + \ldots$$

Anche in caso di serie, nelle quali la regolarità degli intervalli sia interrotta da lacune di osservazione, conviene rendere possibile l'impiego di detta formola, in vista della sua comodità pel calcolo, completando o aggiustando la serie per via di interpolazioni parziali.

La variabile ψ_I è semplicemente l'intervallo in unità di tempo, che separa i singoli termini dal punto di mezzo della serie. Se questa va, poniamo, dal 1892 a tutto il 1902, cioè comprende undici anni, il valore di ψ_I corrispondente al 1897, che è l'anno di mezzo, sarà *zero;* i valori corrispondenti al 1896, 1895, 1894... oppure al 1898, 1899, 1900... saranno rispettivamente eguali a — 1, — 2, — 3... ovvero a + 1, + 2, + 3...

Le variabili ψ_{II}, ψ_{III}, ecc., sono speciali funzioni di ψ_I. I valori che assume ψ_{II} non sono altro che i quadrati delle ψ_I diminuiti della propria media aritmetica; e i valori della ψ_{III} non sono che i cubi delle ψ_I diminuiti dei termini in progressione aritmetica, che si ottengono interpolando una retta fra essi cubi, ecc.

Pertanto convien ricorrere a prontuari, come quello recente del Pareto (1). Noi riporteremo da quest'ultimo, alquanto modificati nella forma, alcuni prospetti per serie da cinque fino a sedici termini, limitatamente al bisogno dell'interpolazione d'una curva di quarto grado.

Facile è la determinazione delle costanti A, B, C ... Si calcoli anzitutto la media aritmetica delle quantità osservate; questa media è il valore cercato di A. Indi si determini B moltiplicando i dati dell'osservazione per i valori di ψ_I; e fatta la somma dei prodotti, la si divida per la somma dei quadrati di ψ_I; il quoziente ottenuto è il valore di B. Similmente si determini C moltiplicando i dati della serie per i valori di ψ_{II} e dividendo la somma dei prodotti per la somma dei quadrati di ψ_{II}; e così di seguito.

Per ciò che abbiam detto testè delle variabili ψ_{II}, ψ_{III}, ecc., come funzioni particolari di ψ_I, si comprende che la formola interpolatoria, di cui si tratta, può presentarsi anche come sviluppo ordinato, secondo le potenze intere di ψ_I, ossia:

$$y = a + b\psi_I + c\psi_I^2 + d\psi_I^3 + \dots$$

(1) Veggasi V. PARETO, *Tables pour faciliter l'application de la méthode des moindres carrés.* Communication présentée à l'Assemblée annuelle des Statisticiens officiels et de la Société suisse de Statistique, tenue à Lausanne en 1898.

Tabella di valori per l'applicazione della formola interpolatoria

$$y = A + B\psi_I + C\psi_{II} + D\psi_{III} + \dots$$

(Il numero dei termini della serie da interpolare è espresso da n).

Per $n = 5$

ψ_I	ψ_{II}	ψ_{III}	ψ_{IV}
− 2	+ 2	− 1,2	+ 24/70
− 1	− 1	+ 2,4	− 96/70
0	− 2	0	+ 144/70
+ 1	− 1	− 2,4	− 96/70
+ 2	+ 2	+ 1,2	+ 24/70

$\Sigma(\psi^2_I) = 10$; $\Sigma(\psi^2_{II}) = 14$;
$\Sigma(\psi^2_{III}) = 14,4$; $\Sigma(\psi^2_{IV}) = 8,16/70$

Per $n = 6$

ψ_I	ψ_{II}	ψ_{III}	ψ_{IV}
− 2,5	+ 10/8	− 3	+ 12/7
− 1,5	− 2/8	+ 4,2	− 36/7
− 0,5	− 8/8	+ 2,4	+ 24/7
+ 0,5	− 8/8	− 2,4	+ 24/7
+ 1,5	− 2/8	− 4,2	− 36/7
+ 2,5	+ 10/8	+ 3	+ 12/7

$\Sigma(\psi^2_I) = 17,5$; $\Sigma(\psi^2_{II}) = 37,1/8$;
$\Sigma(\psi^2_{III}) = 64,8$; $\Sigma(\psi^2_{IV}) = 82,2/7$

Per $n = 7$

ψ_I	ψ_{II}	ψ_{III}	ψ_{IV}
− 3	+ 5	− 6	+ 5,1/7
− 2	0	+ 6	− 12
− 1	− 3	+ 6	+ 1,5/7
0	− 4	0	+ 10,2/7
+ 1	− 3	− 6	+ 1,5/7
+ 2	0	− 6	− 12
+ 3	+ 5	+ 6	+ 5,1/7

$\Sigma(\psi^2_I) = 28$; $\Sigma(\psi^2_{II}) = 84$;
$\Sigma(\psi^2_{III}) = 216$; $\Sigma(\psi^2_{IV}) = 452,4/7$

Per $n = 8$

ψ_I	ψ_{II}	ψ_{III}	ψ_{IV}
− 3,5	+ 7	− 10,5	+ 12
− 2,5	+ 1	+ 7,5	− 22,2/7
− 1,5	− 3	+ 10,5	− 5,1/7
− 0,5	− 5	+ 4,5	+ 15,3/7
+ 0,5	− 5	− 4,5	+ 15,3/7
+ 1,5	− 3	− 10,5	− 5,1/7
+ 2,5	+ 1	− 7,5	− 22,2/7
+ 3,5	+ 7	+ 10,5	+ 12

$\Sigma(\psi^2_I) = 42$; $\Sigma(\psi^2_{II}) = 168$;
$\Sigma(\psi^2_{III}) = 594$; $\Sigma(\psi^2_{IV}) = 1.810,2/7$

Per $n = 9$

ψ_I	ψ_{II}	ψ_{III}	ψ_{IV}
− 4	+ 9,1/8	− 16,8	+ 24
− 3	+ 2,1/8	+ 8,4	− 36
− 2	− 2,2/8	+ 15,6	− 18,6/7
− 1	− 5,2/8	+ 10,8	+ 15,3/7
0	− 6,2/8	0	+ 30,6/7
...	...	...	...

$\Sigma(\psi^2_I) = 60$; $\Sigma(\psi^2_{II}) = 308$;
$\Sigma(\psi^2_{III}) = 1.425,6$; $\Sigma(\psi^2_{IV})$ 5.883,3/7

Per $n = 10$

ψ_I	ψ_{II}	ψ_{III}	ψ_{IV}
− 4,5	+ 12	− 25,2	+ 43,2
− 3,5	+ 4	+ 8,4	− 52,8
− 2,5	− 2	+ 21	− 40,8
− 1,5	− 6	+ 18,6	+ 7,2
− 0,5	− 8	+ 7,2	+ 43,2
...	...	...	...

$\Sigma(\psi^2_I) = 82,5$; $\Sigma(\psi^2_{II}) = 528$:
$\Sigma(\psi^2_{III}) = 3.088,8$; $\Sigma(\psi^2_{IV}) = 16.473,6$

(*Segue*) TABELLA DI VALORI, ECC.

Per $n = 11$

ψ_I	ψ_{II}	ψ_{III}	ψ_{IV}
— 5	+ 15	— 36	+ 72
— 4	+ 6	+ 7,2	— 72
— 3	— 1	+ 26,4	— 72
— 2	— 6	+ 27,6	— 12
— 1	— 9	+ 16,8	+ 48
0	— 10	0	+ 72
...	...	...	...

$\Sigma(\psi^2_I) = 110$; $\Sigma(\psi^2_{II}) = 858$;
$\Sigma(\psi^2_{III}) = 6.177,6$; $\Sigma(\psi^2_{IV})$ 41.184

Per $n = 12$

ψ_I	ψ_{II}	ψ_{III}	ψ_{IV}
— 5,5	$+ 18,{}^1/_3$	— 49,5	$+ 113,{}^1/_7$
— 4,5	$+ 8,{}^1/_3$	+ 4,5	$- 92,{}^4/_7$
— 3,5	$+ {}^1/_3$	+ 31,5	$- 113,{}^1/_7$
— 2,5	$- 5,{}^2/_3$	+ 37,5	$- 44,{}^4/_7$
— 1,5	$- 9,{}^2/_3$	+ 28,5	$+ 41,{}^1/_7$
— 0,5	$- 11,{}^2/_3$	+ 10,5	+ 96
...	...	...	...

$\Sigma(\psi^2_I) = 143$; $\Sigma(\psi^2_{II}) = 1.334,{}^2/_3$;
$\Sigma(\psi^2_{III}) = 11.583$; $\Sigma(\psi^2_{IV}) = 94.134,{}^6/_7$

Per $n = 13$

ψ_I	ψ_{II}	ψ_{III}	ψ_{IV}
— 6	+ 22	— 66	$+ 169,{}^5/_7$
— 5	+ 11	0	$- 113,{}^1/_7$
— 4	+ 2	+ 36	$- 164,{}^4/_7$
— 3	— 5	+ 48	$- 92,{}^4/_7$
— 2	— 10	+ 42	$+ 18,{}^6/_7$
— 1	— 13	+ 24	$+ 109,{}^5/_7$
0	— 14	0	+ 144
...	...	...	...

$\Sigma(\psi^2_I) = 182$; $\Sigma(\psi^2_{II}) = 2.002$
$\Sigma(\psi^2_{III}) = 20.592$; $\Sigma(\psi^2_{IV}) = 200.036,{}^4/_7$

Per $n = 14$

ψ_I	ψ_{II}	ψ_{III}	ψ_{IV}
— 6,5	+ 26	— 85,8	$+ 245,{}^1/_7$
— 5,5	+ 14	— 6,6	— 132
— 4,5	+ 4	+ 39,6	$- 226,{}^2/_7$
— 3,5	— 4	+ 58,8	$- 157,{}^5/_7$
— 2,5	— 10	+ 57,0	$- 22,{}^2/_7$
— 1,5	— 14	+ 40,2	+ 108
— 0,5	— 16	+ 14,4	$+ 185,{}^1/_7$
...	...	...	...

$\Sigma(\psi^2_I) = 227,5$; $\Sigma(\psi^2_{II}) = 2.912$;
$\Sigma(\psi^2_{III}) = 35.006,4$ $\Sigma(\psi^2_{IV}) = 400.073,{}^1/_7$

Per $n = 15$

ψ_I	ψ_{II}	ψ_{III}	ψ_{IV}
— 7	$+ 30,{}^1/_3$	— 109,2	+ 343,2
— 6	$+ 17,{}^1/_3$	— 15,6	$- 147,{}^6/_{70}$
— 5	$+ 6,{}^1/_3$	+ 42	$- 297,{}^{66}/_{70}$
— 4	$- 2,{}^2/_3$	+ 69,6	$- 241,{}^{26}/_{70}$
— 3	$- 9,{}^2/_3$	+ 73,2	$- 85,{}^{26}/_{70}$
— 2	$- 14,{}^2/_3$	+ 58,8	$+ 86,{}^4/_{70}$
— 1	$- 17,{}^2/_3$	+ 32,4	$+ 212,{}^{64}/_{70}$
0	$- 18,{}^2/_3$	+ 0	+ 259,2
...	...	...	...

$\Sigma(\psi^2_I) = 280$; $\Sigma(\psi^2_{II}) = 4.125,{}^1/_3$
$\Sigma(\psi^2_{III}) = 57.283,2$; $\Sigma(\psi^2_{IV}) = 760.138,{}^6/_7$

Per $n = 16$

ψ_I	ψ_{II}	ψ_{III}	ψ_{IV}
— 7,5	+ 35	— 136,5	+ 468
— 6,5	+ 21	— 27,3	— 156
— 5,5	+ 9	+ 42,9	$- 378,{}^6/_7$
— 4,5	— 1	+ 80,1	$- 344,{}^4/_7$
— 3,5	— 9	+ 90,3	$- 173,{}^1/_7$
— 2,5	— 15	+ 79,5	$+ 39,{}^3/_7$
— 1,5	— 19	+ 53,7	$+ 221,{}^1/_7$
— 0,5	— 21	+ 18,9	+ 324
...	...	...	...

$\Sigma(\psi^2_I) = 340$; $\Sigma(\psi^2_{II}) = 5.712$;
$\Sigma(\psi^2_{III}) = 90.698,4$; $\Sigma(\psi^2_{IV}) = 1.382.071$

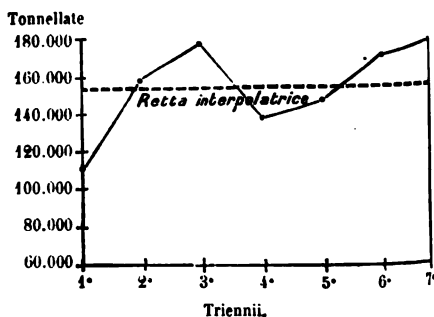
N.B. — Dalla serie di 9 termini in poi si è ommessa per brevità la seconda parte della serie, nella quale si ripetono simmetricamente i termini della prima parte con eguali segni per i valori di ψ_{II} e ψ_{IV} e con segni contrari per i valori di ψ_I e ψ_{III}.

Un facile esame dei prospetti mostra come le variabili ψ_{I}, ψ_{II}, ψ_{III}... costituiscono esse medesime, rispettivamente, una serie lineare o una parabolica di secondo grado o una parabolica di terzo, ecc.; in altre parole, serie tali da avere, rispettivamente, costanti le *differenze prime* dei loro termini, o costanti le *differenze seconde*, o costanti le *differenze terze*, ecc. La somma algebrica per ciascuna serie è sempre zero.

Sia dunque da interpolare la serie seguente, che riguarda la *produzione del ferro* nelle officine italiane dal 1882 al 1902:

Periodi	Produzione media annua in tonnellate
1882-1884	112.080
1885-1887	158.400
1888-1890	178.256
1891-1893	138.329
1894-1896	148.515
1897-1899	171.725
1900-1902	178.101

Una prima e più semplice operazione consisterà nell'interpolare una *retta* parallela all'asse delle ascisse (asse dei tempi), ossia nel sostituire alla media reale di ogni triennio la media aritmetica dell'intero periodo considerato. Si avrà:

Serie osservata	Serie di 1ª interpolazione	Differenze
112.080	155.058	— 42.978
158.400	155.058	+ 3.342
178.256	155.058	+ 23.198
138.329	155.058	— 16.729
148.515	155.058	— 6.543
171.725	155.058	+ 16.667
178.101	155.058	+ 23.043

L'equazione della retta interpolatrice è: $y = 155.058$. Ma gli scarti fra i valori teorici e quelli osservati sono troppo grandi. Occorre procedere ad una seconda interpolazione.

Infatti una maggiore approssimazione si consegue interpolando una retta inclinata sull'asse dei tempi, ossia sostituendo alla serie osservata una serie di numeri teorici formanti una progressione aritmetica, bene inteso alla condizione solita che la somma dei quadrati dei loro scostamenti dai numeri osservati sia un *minimum* in confronto della somma che si otterrebbe adottando una qualsivoglia altra progressione aritme-

tica. Per calcolare una siffatta serie teorica si moltiplicano i numeri dati per ψ_1, i cui valori, come si legge nel prontuario, sono:

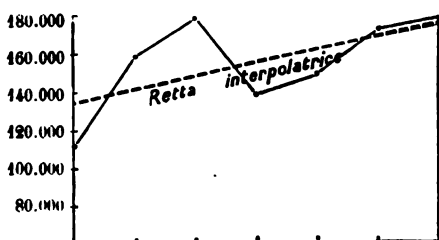
$$-3, -2, -1, 0, +1, +2, +3;$$

poi si fa la somma dei prodotti e la si divide per la somma dei quadrati degli ψ_1, che è 28. Si avrà:

$-3 \times 112.080 = -336.240$		
$-2 \times 158.400 = -316.800$		Coefficiente di ψ_1
$-1 \times 178.256 = -178.256$		
$0 \times 138.329 = 0$		194.972 : 28 = 6.963,3
$+1 \times 148.515 = +148.515$		
$+2 \times 171.725 = +343.450$		
$+3 \times 178.101 = +534.303$		
Somma $= +194.972$		

Si tolga ora o si aggiunga (secondo i segni) alla media aritmetica 155.058, che rappresenta la prima e più rudimentale interpolazione, tante volte la cifra 6963,3, quante sono le unità di ψ_1. Avremo:

Serie della 1ª interp.	Valori di ψ_1			Serie della 2ª interpolazione	Serie osservata	Differenze
155.058	— 3	× 6.963,3	=	134.168,1	112.080	— 22.088,1
155.058	— 2	× 6.963,3	=	141.131,4	158.400	+ 17.268,6
155.058	— 1	× 6.963,3	=	148.094,7	178.256	+ 30.161,3
155.058	± 0	× 6.963,3	=	155.058	138.329	— 16.729
155.058	+ 1	× 6.963,3	=	162.021,3	148.515	— 13.506,3
155.058	+ 2	× 6.963,3	=	168.984,6	171.725	+ 2.740,4
155.058	+ 3	× 6.963,3	=	175.947,9	178.101	+ 2.153,1

La nuova serie teorica (descritta nel diagramma dalla retta punteggiata) è già più approssimata della precedente; gli scarti estremi,

che andavano da — 42.978 a + 23.198, ora non vanno che da — 22.088 a + 30.161. Lo scostamento quadratico medio da 22.449 discende a 17.608. L'equazione della nuova retta è:

$$y = 155.058 + 6.963,3\,\psi_1.$$

E infatti, dando a ψ_1 i valori indicati nel prontuario per una serie di sette termini, si riproducono i numeri della progressione aritmetica, che costituisce la seconda interpolazione.

Un terzo passo innanzi si può tentare interpolando una parabola ordinaria o di secondo grado. All'uopo si moltiplicano le cifre della serie osservata per i valori di ψ_{II}, che nel prontuario sono indicati così:

$$+5,\ 0,\ -3,\ -4,\ -3,\ 0,\ +5;$$

indi si fa la somma dei prodotti e la si divide per la somma dei quadrati degli ψ_{II}, cioè per 84. Avremo dunque:

				Coefficiente di ψ_{II}
+ 5 × 112.080	=	+ 560.400		
0 × 158.400	=	0		
— 3 × 178.256	=	— 534.768		
— 4 × 138.329	=	— 553.316	— 82.724 : 84 =	— 984,8
— 3 × 148.515	=	— 445.545		
0 × 171.725	=	0		
+ 5 × 178.101	=	+ 890.505		
Somma	=	— 82.724		

Ciò fatto, ad ogni termine della seconda interpolazione si aggiunge o si toglie (secondo i segni) tante volte la cifra — 984,8 quante sono le unità degli ψ_{II}:

Serie della 2ª interp.	Valori di ψ_{II}		Serie della 3ª interpolazione	Serie osservata	Differenze
134.168,1	+ 5	× (— 984,8) =	129.244,1	112.080	— 17,164,1
141.131,4	± 0	× (— 984,8) =	141.131,4	158.400	+ 17,268,6
148.094,7	— 3	× (— 984,8) =	151.049,1	178.256	+ 27,206,9
155.058	— 4	× (— 984,8) =	158.997,2	138.329	— 20,668,2
162.021,3	— 3	× (— 984,8) =	164.975,7	148.515	— 16,460,7
168.984,6	± 0	× (— 984,8) =	168.984,6	171.725	+ 2,740,4
175.947,9	+ 5	× (— 984,8) =	171.023,9	178.101	+ 7,077,1

È facile verificare che l'interpolazione della parabola di secondo grado non ci ha fatto fare che un ben piccolo guadagno di approssimazione; lo scostamento (quadratico) medio, che poco fa era di 17.608, lo troveremo ora ridotto appena a 17.274. Anzi, se in luogo dello scostamento quadratico medio avessimo preso a considerare lo scostamento semplice medio, l'approssimazione risulterebbe aver fatto un passo addietro. Infatti, questo scostamento semplice da 14.949, quant'era per la retta, risale a 15.512 per la parabola. Ad ogni modo rinunziando a rappresentare in diagramma la nuova serie, ci basterà dire che la parabola calcolata vi figurerebbe con una lieve curvatura. Tentiamo dunque senz'altro l'interpolazione di una curva di terzo grado.

All'uopo moltiplicheremo le cifre della serie osservata per i valori di ψ_{III}, che nella tabella sono indicati così:

$$-6,\ +6,\ +6,\ 0,\ -6,\ -6,\ +6;$$

e come al solito, fatta la somma dei prodotti, divideremo il risultato per la somma dei quadrati degli ψ_{III}, ossia per 216.

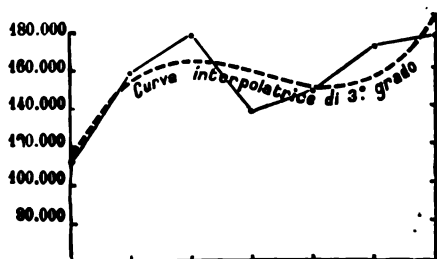
— 6	×	112.080	=	—	672.480	
+ 6	×	158.400	=	+	950.400	Coefficiente di ψ_{III}
+ 6	×	178.256	=	+	1.069.536	
0	×	138.329	=		0	494.622 : 216 = 2.289,92
— 6	×	148.515	=	—	891.090	
— 6	×	171.725	=	—	1.030.350	
+ 6	×	178.101	=	+	1.068.606	
		Somma	=	+	494.622	

Non resta che aggiungere o togliere, secondo i segni, alle cifre della terza interpolazione tante volte il coefficiente 2,289,92 quante sono le unità degli ψ_{III}:

Serie della 3ª interp.	Valori di ψ_{III}				Serie della 4ª interpolazione	Serie osservata	Differenze
129.244,1	— 6	×	2.289,92	=	115.504,6	112.080	— 3.424,6
141.131,4	+ 6	×	2.289,92	=	154.870,9	158.400	+ 3.529,1
151.049,1	+ 6	×	2.289,92	=	164.788,6	178.256	+ 13.467,4
158.997,2	± 0	×	2.289,92	=	158.997,2	138.329	— 20.668,2
164.975,7	— 6	×	2.289,92	=	151.236,2	148.515	— 2.721,2
168.984,6	— 6	×	2.289,92	=	155.245,1	171.725	+ 16.479,9
171.023,9	+ 6	×	2.289,92	=	184.763,4	178.101	— 6.662,4

Il guadagno d'approssimazione è ora abbastanza notevole; lo scostamento medio da 17.274, qual'era alla terza interpolazione, è disceso a 11.687.

L'equazione della curva interpolatrice è:

$$y = 155.058 + 6.963{,}3\,\psi_I - 984{,}8\,\psi_{II} + 2.289{,}9\,\psi_{III}.$$

Non crediamo necessario di spingere oltre l'approssimazione, interpolando una curva di quarto grado; il procedimento sarebbe analogo affatto a quello fin qui seguito. Piuttosto, giusta quanto abbiamo detto delle variabili ψ_{II}, ψ_{III} ..., che sono funzioni particolari di ψ_I, osserveremo che l'equazione

$$y = A + B\psi_I + C\psi_{II} + D\psi_{III} + \dots$$

può mettersi convenientemente sotto forma di sviluppo secondo le potenze intere di ψ_I, ossia, per usare la notazione più comune, di x:

$$y = a + bx + cx^2 + dx^3 + \dots$$

essendo:

$$\psi_I = x;\quad \psi_{II} = x^2 - \frac{n^2-1}{12};\quad \psi_{III} = x^3 - \frac{3n^2-7}{20}x,\ \text{ecc.},$$

dove n indica ancora il numero dei termini della serie. Questo numero essendo nel nostro caso 7, le formole date diventano:

$$\psi_I = x;\quad \psi_{II} = x^2 - 4;\quad \psi_{III} = x^3 - 7x.$$

Pertanto l'equazione:

$$y = 155.058 + 6.963{,}3\,\psi_I - 984{,}8\,\psi_{II} + 2{,}289{,}9\,\psi_{III}$$

si trasforma in quest'altra:

$$y = 155.058 + 6.963{,}3\,x - 984{,}8\,(x^2 - 4) + 2.289{,}9\,(x^3 - 7x)$$

e risolvendo le parentesi e semplificando:

$$y = 158.997{,}2 - 9.066{,}1\,x - 984{,}8\,x^2 + 2.289{,}9\,x^3.$$

B) *Interpolazione per mezzo di funzioni trigonometriche.*

Sommario: § 1. L'interpolazione di serie periodiche mediante le funzioni *seno* e *coseno*. — § 2. Casi di serie miste.

§ **1.** Per i fenomeni periodici, che presentano cioè massimi e minimi ripetentisi a intervalli regolari, conviene l'interpolazione per mezzo di funzioni trigonometriche. Il procedimento è simile a quello adottato per le funzioni algebriche intere, anzi, per quanto concerne la materialità delle operazioni, più semplice e spedito.

Supponiamo un fenomeno statistico il cui periodo sia diviso in dodici unità di tempo o dodici momenti di osservazione separati da eguali intervalli (es.: i dodici mesi dell'anno o le dodici ore pari (o dispari) della giornata). Dividansi idealmente i 360° della circonferenza in dodici parti ognuna dell'ampiezza di 30 gradi. La prima unità di tempo sarà indicata sull'asse delle ascisse da 0°; la seconda da 30°, la terza da 60°, e così via sino alla dodicesima, cui corrisponderanno 330 gradi. Se il ciclo consta di 24 unità di tempo (come le 24 ore della giornata o le 24 quindicine di un anno) si immaginerà la circonferenza divisa in 24 parti, procedenti di 15 in 15 gradi; se consta di sette unità (esempio: i sette giorni della settimana), le divisioni avranno ognuna l'ampiezza di 51°,26'.

Ciò posto, per sostituire alla serie osservata una serie teorica più semplice e regolare esprimibile coll'equazione:

$$y = A + B\ sen\ \varphi$$

col metodo dei minimi quadrati, la via pratica da seguire è facilissima.

Si faccia anzitutto la media aritmetica dei dati forniti dall'osservazione; questa media è il valore di A nell'equazione in parola. Indi si determini il coefficiente B, moltiplicando i dati dell'osservazione per i valori scelti della variabile *sen* φ, sommando tali prodotti e dividendo il risultato per la somma dei quadrati di *sen* φ. Il quoziente fornisce il valore cercato di B.

Abbiasi la serie della *variazione barometrica diurna*, secondo le osservazioni fatte dal Carlini nelle estati del 1826-1830. Per semplicità di calcoli trascuriamo la cifra costante di 751 millimetri, tenendo conto solo dei millimetri e frazioni di millimetro osservati in più della cifra predetta. Il calcolo va disposto così:

Ore di osservazione	Variabile φ espressa in gradi	Altezze barometriche osservate in più di mm. 751	Valori di *sen* φ	Prodotto delle altezze per i valori di *sen* φ
0	0 gradi	mm. 1,095	0	0
2	30 »	» 0,686	0,5	0,343
4	60 »	» 0,229	0,866	0,198
6	90 »	» 0,057	1	0,057
8	120 »	» 0,337	0,866	0,292
10	150 »	» 0,658	0,5	0,329
12	180 »	» 0,822	0	0
14	210 »	» 0,835	— 0,5	— 0,417
16	240 »	» 0,957	— 0,866	— 0,829
18	270 »	» 1,124	— 1	— 1,124
20	300 »	» 1,376	— 0,866	— 1,192
22	330 »	» 1,342	— 0,5	— 0,671
	Media	mm. 0,793	*Somma*	= — 3,014

Dividendo ora la cifra — 3,014 per la somma dei quadrati di *sen* φ, cioè per 6, si ottiene il quoziente — 0,502. Mettendo in conto i mm. 751 trascurati, l'equazione della curva risulta:

$$y = 751{,}793 - 0{,}502 \; sen \, \varphi.$$

Da questa equazione, qualora si diano a *sen* φ i successivi valori 0, 0,5, 0,866, ecc., si ricava una serie teorica abbastanza concordante colla serie osservata. Tuttavia, per un'approssimazione maggiore, conviene interpolare anche la funzione *coseno*, mediante un procedimento analogo a quello ora seguito, cioè col moltiplicare i dati dell'osservazione per i valori di *cos* φ (1), sommare i prodotti e dividere il risultato

(1) Questi valori sono, a partire da 0° e progredendo di 30 in 30 gradi. i seguenti: 1; 0,866; 0,5; 0; — 0,5; — 0,866; — 1; — 0,866; — 0,5; 0; 0,5; 0,866.

per la somma dei quadrati dei *cos* φ, che è ancora 6. Il quoziente 0,149, fornisce il coefficiente cercato di cos φ, ossia l'equazione si scriverà così:

$$y = 751,793 - 0,502 \text{ sen } \varphi + 0,149 \cos \varphi.$$

Da essa si ricavano valori maggiormente approssimati a quelli della osservazione. Se ancora si volesse guadagnare in approssimazione, converrebbe interpolare la funzione *sen* 2 φ (seno del doppio dell'angolo) e poi quella di *cos* 2 φ (coseno del doppio dell'angolo) (1), potendo darsi che il fenomeno periodico in esame, insieme ad un ciclo principale, noverasse due cicli secondari di mezzo periodo ciascuno (nel caso concreto, di mezza giornata ciascuno). Così si può procedere a interpolare la funzione *sen* 3 φ e *cos* 3 φ per mettere eventualmente in risalto un ciclo triplo della durata di otto ore. Spinta sufficientemente innanzi, la formola interpolatrice riprodurrebbe esattamente i dati dell'osservazione; ma allora il vantaggio della semplicità svanisce, la formola restando ingombrata da termini, che riflettono le ondulazioni *parassite* (come le chiama lo Schiaparelli) della serie, mascheranti la vera legge del fenomeno.

Arrestando l'interpolazione al termine *cos* 2 φ, si ha:

$$y = 751,793 - 0,502 \, sen \, \varphi + 0,149 \cos \varphi$$
$$- 0,145 \, sen \, 2\varphi + 0,174 \cos 2\varphi \quad (2).$$

(1) Cioè si dovranno moltiplicare i dati dell'osservazione per i valori qui indicati nella 3ª e 4ª colonna e fatta la somma dei prodotti dividerla per la somma dei quadrati di *sen* 2 φ o *cos* 2 φ, che è ancora = 6.

φ	2 φ	*Sen* 2 φ	*Cos* 2 φ
0 gradi	0 gradi	0	1
30 »	60 »	0,866	0,5
60 »	120 »	0,866	— 0,5
90 »	180 »	0	— 1
120 »	240 »	— 0,866	— 0,5
150 »	300 »	— 0,866	0,5
180 »	360 = 0 »	0	1
210 »	420 = 60 »	0,866	0,5
240 »	480 = 120 »	0,866	— 0,5
270 »	540 = 180 »	0	— 1
300 »	600 = 240 »	— 0,866	— 0,5
330 »	660 = 300 »	— 0,866	0,5

(2) Tale formola può anche scriversi così:

$$y = 751,793 + 0,524 \text{ sen } (\varphi + 163^\circ,28')$$
$$+ 0,227 \text{ sen } (2\varphi + 129^\circ,48')$$

Infatti, prendendo due valori qualsiasi per l'angolo φ, come 0° e 30°, si avrà dalla formola data nel testo:

Per $\varphi = 0^\circ$ $\quad -502 \text{ sen } 0^\circ + 0,149 \cos 0^\circ = 0,149$

Per $\varphi = 30^\circ$ $\quad -502 \text{ sen } 30^\circ + 0,149 \cos 30^\circ = -0,122$

Queste espressioni si vogliono trasformare in due altre, in cui figuri un

Applicando la qual formola e confrontandone i valori teorici con quelli della serie empirica data, otteniamo:

Ore di osservazione	Altezze medie osservate	Altezze calcolate colla formola precedente	Differenze tra il calcolo e l'osservazione
0	752,095	752,116	— 0,021
2	751,686	751,633	+ 0,053
4	751,229	751,220	+ 0,009
6	751,057	751,117	— 0,060
8	751,337	751,322	+ 0,015
10	751,658	751,626	+ 0,032
12	751,822	751,819	+ 0,003
14	751,835	751,877	— 0,042
16	751,957	751,941	+ 0,016
18	752,124	752,121	+ 0,003
20	752,376	752,341	+ 0,035
22	752,342	752,386	— 0,044

---

coefficiente unico x moltiplicato per il *seno* della variabile φ aumentata di una costante k, cioè:

$$x \text{ sen } (0^{\circ} + k) = 0{,}149$$

(che può scriversi $x \text{ sen } k = 0{,}149$)

$$x \text{ sen } (30^{\circ} + k) = -0{,}122.$$

Dividendo la prima per la seconda e semplificando si avrà:

$$\frac{\text{sen } k}{\text{sen } (30^{\circ} + k)} = -1{,}221.$$

Ora

$$\text{sen } (30^{\circ} + k) = \text{sen } 30^{\circ} \cos k + \text{sen } k \cos 30^{\circ};$$

perciò, essendo

$$\text{sen } 30^{\circ} = 0{,}5; \quad \cos 30^{\circ} = 0{,}866;$$

con brevi operazioni arriveremo all'espressione:

$$\frac{\text{sen } k}{\cos k} = \frac{-0{,}6105}{2{,}0574},$$

ossia

$$\text{tang } k = -0{,}29673,$$

donde

$$k = 180^{\circ} - 16^{\circ},31',38'' = 163^{\circ},28',22''.$$

Conosciuto k, è facile conoscere x, utilizzando la formola data sopra:

$$x \text{ sen } k = 0{,}149.$$

Si avrà in altri termini:

$$x \text{ sen } 163^{\circ},28',22'' = 0{,}149$$

ed essendo sen $163^{\circ},28',22'' = 0{,}28447$, sarà:

$$x = 0{,}524.$$

Le differenze tra il calcolo e l'osservazione sono così piccole da potersi ritenere l'effetto di cause perturbatrici accidentali. Arrestando dunque la formola ai termini in φ e 2φ, concludesi che la variazione barometrica diurna intorno al solstizio estivo consta di un'onda principale ripetentesi ogni 24 ore e di un'onda secondaria della durata di 12 ore, epperò ripetentesi due volte nello spazio di una giornata.

Con questo metodo lo studioso può esercitarsi a trovare l'espressione analitica di molti fenomeni periodici, come la natalità, la mortalità, il consumo di certe derrate, il movimento ferroviario di certe merci, il movimento bancario di depositi, sconti, ecc.

§ 2. Presentandosi il caso di un fenomeno, che abbia al tempo stesso i caratteri della periodicità e la tendenza a seguire una curva non periodica di un dato ordine, converrà interpolare prima una funzione algebrica intera (operando, beninteso, su periodi compiuti e non incompleti) per mettere in evidenza l'andamento generale del fenomeno, astrazion fatta dalla sua periodicità; indi, formata la serie degli scostamenti tra i valori teorici e quelli osservati, interpolarla mediante una funzione trigonometrica.

C) *Critica dei procedimenti d'interpolazione.*

Sommario: § 1. Arbitrio nella scelta della funzione-tipo. — § 2. Osservazioni sul metodo dei minimi quadrati. — § 3. I *fatti nuovi* e le interpolazioni per parti di serie. — § 4. Criteri per la separazione della parte regolare dall'irregolare nelle serie.

§ 1. Da quanto abbiamo detto intorno al terzo ufficio dell'interpolazione s'intravvede già che non esistono criteri assoluti per separare nelle serie empiriche la parte costante o variabile regolare, che esprimerebbe la vera legge del fenomeno, dalla parte variabile senza regola definita, che attesta l'intervento di cause secondarie perturbatrici. E infatti la soluzione del problema presenta molti aspetti arbitrari. Limitare l'arbitrio vorrebbe dire avvicinare per grado di perfezione le leggi empiriche alle leggi naturali esatte, ossia far fare un passo immenso alle scienze o parti di scienze, che si valgono del metodo statistico. Degli aspetti arbitrari dei procedimenti di interpolazione scrisse già con grande competenza lo Schiaparelli (1), e il tema fu ripreso recentemente dal Pareto con particolare riguardo al problema delle *corre-*

---

(1) Veggasi G. V. SCHIAPARELLI, *Sul modo di ricavare la vera espressione delle leggi della natura dalle curve empiriche* (In appendice alle *Effemeridi astronomiche di Milano per l'anno* 1867).

lazioni (1); noi ci proponiamo di renderlo più famigliare, che oggi non sia, agli studiosi di Statistica, mettendoci del nostro qualche non inutile considerazione.

Anzitutto appare arbitraria la scelta della funzione-tipo, nulla giustificando la preferenza data alle funzioni algebriche intere o alle funzioni periodiche procedenti secondo i seni e coseni dell'angolo e dei multipli esatti dell'angolo; nulla, diciamo, salvo la maggiore comodità e speditezza del calcolo. Il che, nota lo Schiaparelli, è un vantaggio subbiettivo del calcolatore; la natura nello stabilire le sue leggi, non si è certo arrestata dinnanzi alla complicatezza delle funzioni.

Tuttavia l'arbitrio è meno grande di quello che a prima giunta parrebbe. Già nella natura fisica dominano i numeri semplici. Le leggi delle vibrazioni delle corde, della caduta dei gravi, della riflessione e rifrazione della luce, ecc.; le leggi delle proporzioni definite, delle proporzioni multiple, della periodicità degli elementi chimici, ecc., stanno ad attestarlo. Ma anche se entriamo in certi ordini di fatti collettivi, in cui si considera come tipica la media dei casi osservati o la classe di maggior frequenza, ci si presentano fenomeni che variano in ragion semplice (diretta o inversa) di altri, o in ragion dei quadrati, delle radici quadrate, ecc. Tali in antropometria le correlazioni tra pesi e stature, tra stature e frequenza del polso, ecc. In demografia, la ripartizione dei redditi figurata a doppia scala logaritmica, dà luogo, come vedemmo, ad una retta; la ripartizione di un gruppo di coetanei per statura, forza, peso, ecc., figurata a scala logaritmica semplice, dà luogo ad una curva parabolica di secondo grado. Cosicchè o immediatamente o mediatamente coll'artificio dei logaritmi ricompaiono le funzioni algebriche intere. La periodicità dei fenomeni sociali è in generale una conseguenza della periodicità di fenomeni cosmici e come questa si dispiega in molti casi per numeri semplici, così quella. Ad esempio, l'anno forma un ciclo unico sotto molti riguardi; ma nel corso dell'anno le posizioni della terra rispetto al sole producenti estremi di temperatura sono due a intervallo di sei mesi (gennaio e luglio) e quelle producenti temperature medie (aprile e ottobre) sono pure due. Di riflesso abbiamo due massimi e due minimi nella mortalità, cioè un'onda doppia associata ad un'onda semplice. Naturale quindi l'interpolazione secondo i seni e coseni dell'angolo e del doppio dell'angolo; mentre non vedremmo ragion sufficiente di procedere secondo le potenze intere dei seni e coseni, sebbene le funzioni $sen^2 \varphi$, $cos^2 \varphi$, siano altrettanto semplici e comode pel calcolo quanto le funzioni $sen\ 2\varphi$ e $cos\ 2\varphi$. Tutto

---

(1) Veggasi V. PARETO, *Quelques exemples d'application des méthodes d'interpolation à la Statistique* (nel *Journal de la Société de Statistique de Paris*, numero di novembre 1897).

considerato dunque, si può affermare che la preferenza data alle funzioni algebriche intere e alle trigonometriche procedenti per multipli dell'angolo, senza essere immune d'arbitrio, soddisfa in generale alla condizione del minimo arbitrio.

§ **2.** Lo stesso metodo dei minimi quadrati non può vantare ragioni assolute in suo favore, che non siano quelle della semplicità dei calcoli. Perchè delle infinite serie lineari o paraboliche, che si possono sostituire alla serie osservata, fermo restando il numero di termini e la somma, noi adottiamo giusto quella che realizza la condizione del minimo nella somma dei quadrati degli scostamenti fra i suoi termini e quelli corrispondenti dell'osservazione? Si dice: per limitare l'arbitrio. Ma l'arbitrio, se non esiste più per i calcolatori *uti singuli*, sussiste per loro *uti universi*, come collettività, in quanto convennero di attenersi a quel principio. La natura dei casi potrebbe suggerire di rendere minima la somma dei cubi (considerati tutti di segno positivo) o delle quarte potenze, ecc.; ovvero di eguagliare il maggiore degli scostamenti semplici positivi al maggiore degli scostamenti semplici negativi. Il metodo dei minimi quadrati, giova ricordarlo, sebbene in grado minore di altri metodi, fa pesare notevolmente nella serie le variazioni eccezionali; obbliga in certo modo la curva interpolatrice a piegarsi verso di esse, sottraendo qualche cosa all'approssimazione di tutti i rimanenti termini. Sia, ad esempio, la serie:

x	y
— 3	53
— 2	39
— 1	52
0	51
+ 1	49
+ 2	49
+ 3	48
Media . . .	48,71

Nel caso qui esposto, la semplice ispezione ad occhio suggerisce che la serie degli y sarebbe bene rappresentata da una retta *discendente* verso l'asse dei tempi (x), anzichè ascendente, la variazione eccezionale del secondo termine, 39, non potendo modificare l'impressione visiva che si ritrae dalla generalità degli altri termini. In quella vece l'equazione della retta interpolata col metodo dei minimi quadrati, che è:

$$y = 48{,}71 + 0{,}07\,x$$

accenna a un movimento piuttosto ascensionale; e ciò appunto a motivo dell'importanza che acquista nel calcolo la depressione del secondo termine della serie. Ma non può essere dubbio che l'impressione visiva sia, nella fattispecie, la più giusta. Infatti se la causa perturbatrice, che

si cela nel secondo termine, è di quelle che si verificano una volta ogni venti o trent'anni e si trova per combinazione inclusa in una serie o in un tratto di serie di soli sette anni, il dato corrispondente non dovrebbe entrare in calcolo se non per quel tanto per cui c'entrano gli estremi soliti a verificarsi in periodi settennali.

In altre parole, se su molte altre serie settennali della stessa specie e ridotte ad egual somma si trovasse che la media dei minimi non si distacca dalla media delle medie che di 6 punti, il termine 39 della serie data dovrebbe entrare in calcolo come se fosse 42,71 (= 48,71 — 6).

Ogni brusca variazione è indizio non dubbio dell'intervento d'una causa nuova, transitoria o no; e, come vedremo ora, quando un fatto nuovo sopravviene a serie incominciata, non è più rigoroso operare una interpolazione unica per tutta la serie, ma si deve procedere per via di interpolazioni parziali nei tratti separati dalla variazione straordinaria.

§ **3.** Questo dei *fatti nuovi* è altro dei punti magistralmente discussi dallo Schiaparelli. In sostanza egli dice che a formare le costanti dell'equazione (i coefficienti della variabile) concorrono e collaborano tutti i termini della serie, laddove siffatta collaborazione o solidarietà spesso non ha ragion d'essere. Quando una causa nuova sopraggiunge a metà o a un terzo della serie, perchè dovremmo supporla operante anche prima del tempo in cui ha cominciato realmente a manifestarsi? Eppure questa supposizione è implicita nel procedimento stesso di calcolo dei coefficienti. Lo Schiaparelli fa l'esempio della variazione diurna della temperatura: « In 24 ore essa passa per quattro stadii differenti, l'uno separato dall'altro per mezzo di una soluzione di continuità, invero non molto apparente, ma tuttavia dimostrabile. Infatti durante la notte non esiste altra causa che quella dei movimenti di calorico nell'atmosfera per irradiazione e per conduttibilità. Ma, al cominciare dell'alba, entra in giuoco una *nuova causa*, la riflessione dei raggi calorifici del sole nell'atmosfera, analoga alla riflessione della luce, che produce i crepuscoli. Finalmente allo spuntare del sole entra in campo l'irradiazione diretta di questo. È evidente dopo ciò che la temperatura diurna non può, in teoria, rappresentarsi nelle sue variazioni orarie con una formola unica valevole per le 24 ore, siccome non può per eguali ragioni rappresentarsi con una formola unica la variazione diurna della quantità d'illuminazione d'un punto esposto a cielo completamente libero. Anche qui occorrono quattro stadii diversi: notte completa, illuminazione solare nel giorno, illuminazione crepuscolare del mattino e della sera ».

Riflessione giustissima, la cui portata è questa: se è noto il momento nel quale il fatto nuovo interviene nella serie, converrà studiarne separatamente l'influenza e, occorrendo, procedereper via d'interpolazioni parziali; se non è noto, dovremo per il postulato del minimo

arbitrio considerare l'influenza di quel fatto diffusa, per così dire, su tutta la serie.

Un unico tracciato continuo per figurare lo sviluppo del nostro commercio colla Francia dal 1882 al 1904 non sarebbe da approvarsi; l'applicazione delle tariffe differenziali nel 1888, la loro sostituzione colla tariffa massima nel 1892 e il trattato del 1899 costituiscono i « fatti nuovi », per dir solo dei principali, che spezzano la serie ed obbligano ad interpolazioni parziali.

Così, per conoscere come probabilmente si ripartirono nel consumo di varî anni le provviste eccezionali di caffè e zucchero, con cui gli speculatori prevennero gli inasprimenti doganali del 1885, faremo una interpolazione; ma non dovremo interpolare la serie *che a cominciare dal* 1885 e non da anni anteriori a questo; altrimenti si verrebbe al risultato assurdo di ripartire su anni già passati l'effetto degli approvvigionamenti straordinari fatti nel 1885, laddove è chiaro che essi saranno entrati nel consumo degli anni *successivi* e non degli anteriori.

Andando al fondo della questione, noi troviamo che i processi interpolatorii traducono in forme geometriche *pure* le forme *miste* ottenibili dalla rappresentazione grafica immediata delle serie statistiche. Una forma geometrica dicesi « pura » quando tutti i suoi punti derivano da una medesima norma di costruzione (esempio: la retta, il circolo, la parabola, ecc.); sì che data una parte, piccola a piacere, purchè finita, il resto della forma è intieramente determinato. Invece un poligono è una forma « mista », perchè i tratti rettilinei, di cui si compone, avendo diverse giaciture e direzioni, i punti dell'uno sono descritti con norma diversa da quelli dell'altro; sicchè, data una porzione piccola a piacere, purchè finita, della figura, il resto non è punto determinato (1).

Orbene, tutte le serie statistiche costituiscono in realtà forme miste, ove i cambiamenti delle norme di costruzione da tratto a tratto corrispondono all'intervento di cause nuove o alla scomparsa di cause, che avevano operato sino a quell'istante; mentre le serie teoriche, derivanti da un processo interpolatorio, costituiscono forme pure, in cui *si presume* tradotta la legge vera del fenomeno, non inquinata da perturbazioni accidentali, nè da sopravvenienze o disparizioni di speciali fattori.

§ **4.** A qual punto della formola porteremo l'interpolazione, affine di distinguere il meglio possibile il costante dall'accidentale? Ecco altra materia di apprezzamenti variabili da individuo ad individuo. Ci arresteremo ai primi due o tre termini della formola analitica? Interpoleremo una retta o una parabola ordinaria o andremo alle curve di terzo, quarto o quinto grado?

(1) Veggasi SCHIAPARELLI, nell'appendice all'opera già citata del VIGNOLI, *Peregrinazioni antropologiche e fisiche*, pag. 273 e seguenti.

Regole, *a priori*, è difficile darne; molto dipende dall'intuito dell'osservatore. Converrà vedere se la grandezza e saltuarietà degli scostamenti positivi e negativi tra i valori forniti dall'osservazione e quelli ottenuti a calcolo a diversi stadii del processo interpolatorio, son tali da giustificare l'imputazione alle sole cause accidentali. Bisognerà por mente ancora ai guadagni ineguali di approssimazione, che si realizzano a diversi stadii del calcolo (1). Si guarderà, infine, se la formola interpolatrice conduce più o meno presto all'assurdo o all'improbabile, quando la si immagini valevole per l'avvenire; chè, se vi conduce lentamente, noi riposiamo nella fiducia che nuovi fatti, qualunque essi siano, interverranno certo a modificare l'andamento della curva; ma, se vi conduce assai presto, dobbiamo indagare quali nuovi fatti sono ragionevolmente da aspettarsi a vicina scadenza, che impediscano la corsa all'impossibile o all'improbabile. Spieghiamoci con un esempio.

Le percentuali degli sposi, che seppero firmare l'atto nuziale in Italia, formano dal 1872 al 1900 due serie rappresentabili approssimativamente colle equazioni:

Per gli sposi

$$y = 56{,}48 + 0{,}758\,x - 0{,}011\,x^2$$

Per le spose

$$y = 35{,}91 + 0{,}995\,x + 0{,}0108\,x^2$$

fissata l'origine al 1886. Supponendo valide le formole per gli anni posteriori al 1900, le percentuali *prevedibili* pel 1901 e 1902 ottenute dando alla x i valori 15 e 16 (perchè il 1901 e il 1902 sono infatti di 15 e di 16 anni, rispettivamente, posteriori al 1886), possono essere messe oggi a confronto colle percentuali risultate dalla osservazione statistica:

SPOSI				SPOSE			
ANNI	Percentuale		Differenze	ANNI	Percentuale		Differenze
	Prevedibile	Osservata			Prevedibile	Osservata	
1901	65,37	67,26	— 1,89	1901	53,25	53,90	— 0,65
1902	65,79	67,44	— 1,65	1902	54,57	54,21	+ 0,36

(1) Così, nella serie trattata a pag. 170 e seguenti il guadagno d'approssimazione conseguito interpolando la parabola di secondo grado, risultò piccolo e limitato allo scostamento quadratico medio; se in luogo di quest'ultimo avessimo preso in considerazione lo scostamento semplice medio, anzi che guadagno avremmo registrato un regresso nell'approssimazione in confronto della interpolazione lineare. Quindi la formola cui siamo là pervenuti potrebbe semplificarsi eliminando il termine — 984,8 ψu.

Fin qui tra la previsione e l'accertamento c'è un discreto accordo. Ma l'improbabile si presenta tosto, se immaginiamo valida la formola per un più lungo tratto di tempo. Infatti per il 1912 dovremmo aspettarci una percentuale di spose, capaci di firmare l'atto nuziale, superiore a quella degli sposi; 69,08 contro 68,75; prima inverosimiglianza, di cui non è difficile intendere le ragioni. In seguito poi, mentre per le prime la curva tende rapida al livello 100 %, suo limite pratico di arresto, per i secondi raggiungerebbe pigramente il suo massimo alla quota 69,50 %, per poi ridiscendere; cosa più inverosimile ancora. Il fatto, che sin da oggi esclude un risultato di questo genere, è la frequenza crescente degli alunni maschi delle scuole elementari, i quali fra vent'anni passeranno in gran parte a matrimonio, ancora capaci di scrivere il proprio nome. Sicchè la portata delle due formole, massime della prima, non deve estendersi che a breve spazio di tempo, fuori dei limiti della osservazione; la qual cosa può ripetersi per tutte le altre espressioni analitiche di serie empiriche (1).

Ogni interpolazione vuol essere preceduta da un lavoro preparatorio inteso a depurare la serie dalle variazioni dovute a cause speciali conoscibili. Così sapendosi che in febbraio si anticipano molte nozze che, senza il divieto della Quaresima, si celebrerebbero in marzo, e che in novembre se ne anticipano, a danno del dicembre, causa l'Avvento, converrà perequare i mesi così perturbati ed operare una interpolazione parziale, e indi sulla serie così aggiustata procedere all'interpolazione generale. Altrimenti l'efficacia di queste cause, che notoriamente non oltrepassa certi limiti di tempo, si fingerebbe diffusa sull'intero periodo considerato.

Analogamente, quando risultasse che il fenomeno in esame obbedisce nelle sue variazioni alle variazioni di un altro, sarebbe opportuno eliminare dalla serie data la quantità d'effetto imputabile approssimativamente al fenomeno perturbatore, il che si ottiene mediante un calcolo di *correlazione*, di cui diremo tra breve, e la serie così depurata si assoggetterà ad interpolazione.

Appartiene allo stadio preparatorio anche l'indagine se il fenomeno in esame possa decomporsi in gruppi scelti aventi una diversa legge di

(1) Nella previsione delle entrate di bilancio, gli scrittori di finanza e i pratici seguono ancora metodi semplicisti, in confronto dei quali i procedimenti d'interpolazione rappresenterebbero un progresso notevole. Infatti si usa o inscrivere tali e quali per l'esercizio venturo gli accertamenti dell'ultimo esercizio conosciuto (semprechè non sia intervenuta una legge a variare i contingenti d'imposta, le aliquote o le tariffe), o accrescere le risultanze dell'ultimo esercizio di una quantità eguale all'incremento medio annuo osservato, poniamo, nell'ultimo quinquennio. Questo secondo metodo, detto delle *plusvalenze*, è meno semplicista del primo, ma potrebbe essere sostituito con vantaggio da una vera e propria interpolazione estesa a tutto il tratto di serie, che non appare perturbato da mutamenti legislativi o da altri fatti di importanza eccezionale.

sviluppo o di periodicità. Le regolarità statistiche non sono infatti che interferenze più o meno complicate di leggi empiriche.

Trattando più tardi della statistica considerata come forma di induzione, faremo posto al quesito, veramente interessante, se i singoli termini di una formola analitica segnalino la presenza di cause distinte o di gruppi distinti di cause.

D) *Procedimenti interpolatorii per alcune seriazioni particolari.*

Sommario: § 1. Curva dei redditi; valore della costante caratteristica α per diversi paesi, secondo il Pareto. — § 2. Curva dei patrimoni ereditarii in Italia, Francia e Inghilterra. — § 3. Curve del genere binomiale; loro riduzione a parabole logaritmiche.

§ 1. Curve che spesso ricorrono nella rappresentazione a scala naturale di seriazioni riguardanti fenomeni demografici, economici, biologici, ecc., sono le *iperboliche* e le *binomiali*. Facili procedimenti interpolatorii ne dànno, nei casi più semplici, l'espressione analitica.

Sia da interpolare la seriazione dei redditi, di cui a pag. 147, che figurata a doppia scala logaritmica fa luogo, come vedemmo, ad una linea quasi retta. Si prendano i logaritmi di x, ossia dei redditi-limite considerati, e i logaritmi di y, ossia del numero dei contribuenti con reddito superiore ad x; indi fatta la media aritmetica di ciascuna delle due serie, si calcolino le differenze dei singoli termini dalla media rispettiva e, sommate le differenze positive (o negative) della prima serie e le *corrispondenti differenze* della seconda, si divida il secondo risultato per il primo. Il quoziente fornisce il valore di α nell'equazione:

$$\log y = k - \alpha \log x.$$

Ecco, del resto, nel caso pratico, le operazioni:

x	y	$\log x$	$\log y$	Differenze dalla media aritmetica	
				$\Delta \log x$	$\Delta \log y$
1.000	59.486	3	4,7744	— 0,7603	+ 1,0414
2.000	26.968	3,3010	4,4308	— 0,4593	+ 0,6978
4.000	9.766	3,6021	3,9897	— 0,1582	+ 0,2567
7.000	4.264	3,8451	3,6298	+ 0,0848	— 0,1032
10.000	2.397	4	3,3797	+ 0,2397	— 0,3533
15.000	1.310	4,1761	3,1173	+ 0,4158	— 0,6157
25.000	645	4,3979	2,8095	+ 0,6376	— 0,9235
	Medie	3,7603	3,7330	*Somme* — 1,3778 / + 1,3779 (1)	+ 1,9959 / — 1,9957 (2)

(1-2) Queste cifre dovrebbero riuscire, salvo il segno, rigorosamente eguali: il piccolo disaccordo che esiste nell'ultimo decimale dipende dall'aver arrotondati i logaritmi dei numeri dati e le loro medie.

Dividendo ora 1,9959 per — 1,3778 (o dividendo — 1,9957 per 1,3779) si ottiene come quoziente: — 1,45 circa. La qual cosa significa che ognuno dei $\Delta \log y$ è ***in media*** eguale al corrispondente $\Delta \log x$ moltiplicato per — 1,45. Scriviamo dunque:

$$\Delta \log y = -1{,}45\ \Delta \log x.$$

Ma $\Delta \log y = \log y - 3{,}7330$ e similmente $\Delta \log x = \log x - 3{,}7603$; perciò:

$$\log y - 3{,}7330 = -1{,}45\ (\log x - 3{,}7603)$$

e risolvendo:

$$\log y = 9{,}1854 - 1{,}45 \log x$$

equazione di una *retta*, che è l'interpolatrice della linea quasi retta ottenuta a pag. 147 colla rappresentazione a doppia scala logaritmica.

Liberando dai logaritmi tale equazione, si ha:

$$y = \frac{A}{x^{1,45}}$$

equazione dell'*iperbole* interpolatrice della linea, che si sarebbe ottenuta rappresentando a scala naturale la ripartizione dei contribuenti per reddito. La costante A è il numero avente per logaritmo 9,1854.

Verifichiamo la formola. Si cerchi il numero teorico dei contribuenti con reddito superiore a lire 10.000. Si avrà:

$$\log y = 9{,}1854 - 1{,}45 \log 10.000.$$
$$\log y = 3{,}3854.$$

Il numero avente per logaritmo 3,3854, è 2428. Sarebbero cioè 2428 i contribuenti con reddito superiore a 10.000 lire. Il prospetto, esposto sopra, ne indica 2397, ossia soli 31 in meno. Il calcolo e l'osservazione vanno abbastanza d'accordo.

Verifichiamo ancora. Si cerchi il numero teorico di possessori di un reddito superiore a 25.000 lire. Si avrà:

$$\log y = 9{,}1854 - 1{,}45 \log 25.000$$
$$\log y = 2{,}8084.$$

Il numero avente per logaritmo 2,8084 è 643. La nostra tabella dava il numero 645. La differenza tra il calcolo e l'osservazione risulta minima.

Differenze più notevoli si avrebbero per i redditi piccoli; ma qui bisogna ricordare che quanto più discendiamo verso il limite di esenzione dall'imposta, tanto maggiore è il numero di coloro che riescono a sfuggire alla medesima; sicchè forse la curva teorica riesce a correggere in qualche modo le lacune della rilevazione; in secondo luogo che la seriazione dei redditi è di tal natura, che anche figurata a doppia scala logaritmica deve dare in realtà, non una retta, ma una curva sia pure a lievissima inflessione, sicchè per approssimarci di più al vero occorrerebbe un'interpolazione del genere di quella che esporremo al prossimo paragrafo.

Il grande vantaggio del procedimento interpolatorio è anche qui quello di permettere l'inserzione di termini. Se volessimo sapere, ad esempio, quanti dei 59.486 contribuenti, sopra classificati, vivono di un reddito di oltre lire 8300, la risposta non si avrebbe immediata dai dati della osservazione, che passano senz'altro dal limite di 7000 a quello di 10.000 lire; ma si avrebbe facilmente dalla formola:

$$\log y = 9{,}1854 - 1{,}45 \log 8300$$

donde:

$$\log y = 3{,}5027$$

cui corrisponde il numero 3182.

Così se si domandasse, quante persone del gruppo in esame dispongono di un reddito compreso tra 5000 e 6000 lire, basterebbe calcolare nel modo anzidetto quante hanno un reddito superiore a 5000 e quante superiore a 6000 e fare la differenza dei risultati.

Si può anche sconfinare dai limiti della seriazione osservata, fino a porre il quesito: quant'è il reddito spettante a quell'unico contribuente, tra i 59.486, che si eleva più d'ogni altro nella graduatoria di ricchezza? Allora $y = 1$; $\log y = 0$. Si avrà dunque:

$$0 = 9{,}1854 - 1{,}45 \log x$$

donde:

$$\log x = 6{,}33476$$

$$x = 2.161.000 \text{ lire.}$$

Ma è bene avvertire che queste formole valgono per il tratto di curva considerato e non debbono impiegarsi senza riserve al di là dei limiti della serie empirica. Altrimenti si può andare incontro anche a risultati assurdi. Per esempio, ad un reddito *zero* corrisponderebbe un numero infinito di individui.

Con siffatto metodo il Pareto (1) ha posto in chiaro che la ripartizione dei redditi in diversi paesi e tempi obbedisce alla stessa semplice legge; cioè, rappresentata su doppia scala logaritmica dà luogo ad una linea che è per un buon tratto sensibilmente retta. L'inclinazione della retta interpolatrice è data dal coefficiente α, il cui valore oscilla intorno ad 1,50. Quanto maggiore è α, tanto più rapido è il diminuire dei titolari ad ogni gradino della scala del reddito, cioè tanto meno disuguale è la ripartizione della ricchezza. Infatti, se fosse $\alpha = 2$, vorrebbe dire che il numero dei possessori varia in ragione inversa dei quadrati dei redditi-limite e cioè che a un reddito-limite due, tre, quattro volte superiore di un altro corrisponde un numero di titolari 4, 9, 16 volte minore. Se fosse $\alpha = 3$, il numero dei possessori varierebbe in ragione inversa dei cubi dei redditi-limite considerati. Se α fosse

(1) Veggasi *Cours d'Économie politique*, vol. II, libro III, cap. 1°, *La courbe des revenus*, Lausanne, Rouge, 1897.

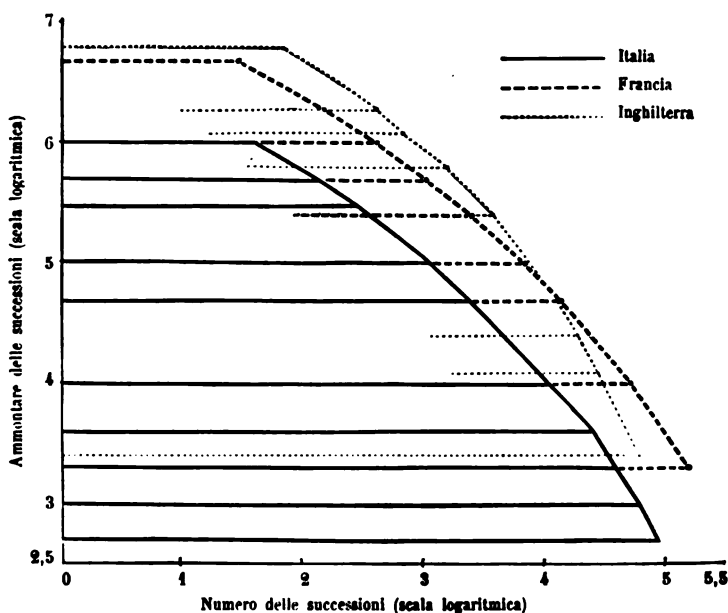
ancor più grande, evidentemente basterebbe salire di pochi gradini la scala del reddito per non trovar più nessun titolare; e se ne dovrebbe dedurre appunto che i diversi strati della popolazione differiscono assai poco tra loro per condizioni economiche.

Paesi	Valori di α
Inghilterra (1879)	1,35
Prussia (1894)	1,60
Sassonia (1886)	1,51
Basilea (1887)	1,24
23 Città italiane (1887)	1,45
Parigi (attualmente)	1,42
» (nel 1292)	1,32
Augsburg (1498)	1,47
Perù (fine XVIII secolo)	1,79

Il prospettino dei valori di α per diversi paesi mostra come considerevoli masse di dati possano venir ridotte ad espressioni algebriche estremamente semplici e reciprocamente confrontate. In ciò sta l'importanza capitale dei procedimenti interpolatorii.

§ 2. Molti fenomeni economici, graduabili per valore, rappresentati in diagrammi a doppia scala logaritmica, non fanno luogo ad una linea che si approssimi ad una retta, come s'è visto nel caso dei redditi, ma a una curva decisamente parabolica. Citiamo la ripartizione per valore delle successioni ereditarie, dei depositi a risparmio, delle cambiali emesse e dei protesti, ecc. Allora l'interpolazione può seguire parecchie vie. Prendiamo, ad esempio, le successioni ereditarie in Italia, Francia e Inghilterra.

ITALIA (1901-1902) SUCCESSIONI		FRANCIA (1902) SUCCESSIONI		INGHILTERRA (1901-1902) SUCCESSIONI	
di ammontare superiore ad x	Numero y	di ammontare superiore ad x	Numero y	di ammontare superiore ad x	Numero y
500 lire	90.100	2.000 fr.	150.234	2.500 fr.	61.393
1.000 »	63.997	10.000 »	52.977	12.500 »	29.080
2.000 »	41.830	50.000 »	13.779	25.000 »	19.612
4.000 »	24.918	100.000 »	6.815	250.000 »	3.839
10.000 »	11.689	250.000 »	2.565	625.000 »	1.593
50.000 »	2.649	500.000 »	1.092	1.250.000 »	685
100.000 »	1.193	1.000.000 »	408	1.875.000 »	413
300.000 »	301	5.000.000 »	27	6.250.000 »	69
500.000 »	137	—	—	—	—
1.000.000 »	41	—	—	—	—

E diamone la rappresentazione grafica a doppia scala logaritmica:

Rappresentazione a doppia scala logaritmica delle successioni in Italia, Francia e Inghilterra.

L'esame di questo diagramma pone in evidenza il parallelismo delle due curve relative alla Francia e all' Italia, cioè dimostra che nei due paesi è press'a poco egualmente rapido il diminuire dei possessori col crescere dei patrimoni posseduti in vita. La diminuzione invece appare più lenta in Inghilterra, prova non dubbia che là è maggiore la disuguaglianza di ripartizione della ricchezza.

L'equazione delle tre curve non può essere, come nel caso dei redditi, $\log y = A - \alpha \log x$, perchè questa è l'equazione di una retta descritta in diagramma a scala logaritmica. L'interpolazione devesi tentare per vie alquanto diverse. Anzitutto si affacciano i due metodi seguenti:

1° Far passare una parabola per tre punti convenientemente scelti, in modo da arrivare all'equazione:

$$\log y = K + m \log x + n (\log x)^2;$$

2° Accrescere la y, oppure la x, di una costante c, in modo da ottenere ancora una retta la cui equazione sia:

$$\log (y + c) = A - \alpha \log x,$$

oppure:

$$\log y = A - \alpha \log (x + c).$$

Senonchè questi metodi hanno l'inconveniente di presentare due costanti caratteristiche (m ed n nel primo caso, α e c nel secondo) vinco-

late colla variabile o vincolate colla variabile e la funzione insieme, la qual cosa rende dubbioso il confronto delle situazioni di due o più paesi diversi (1). *Bisogna trovar modo di ridurre a una sola la costante che caratterizza la seriazione.* Lo scopo si raggiunge *in parte* interpolando i $(\log x)^2$ invece dei semplici $\log x$, cioè determinando l'equazione:

$$\log y = A - \alpha (\log x)^2.$$

Il procedimento è del resto affatto simile a quello già esposto per la ripartizione dei redditi, salvo che invece dei $\log x$ si prendono i loro quadrati. L'elevazione a quadrato dei logaritmi della variabile dà alla linea interpolatrice una curvatura, che l'avvicina alla linea reale del diagramma, quanto si può desiderare per una prima approssimazione.

Ecco d'altronde il procedimento seguito per la seriazione italiana. Non vi facciamo figurare distinte, a causa del loro numero troppo piccolo e troppo variabile d'anno in anno, le successioni superiori a 500.000 lire.

$\log x$	$(\log x)^2$	$\log y$		$\Delta (\log x)^2$	$\Delta \log y$
2,69897	7,28444	4,95472		— 9,36997	+ 0,97660
3	9	4,80616		— 7,65441	+ 0,82804
3,30103	10,89680	4,62149		— 5,75761	+ 0,64337
3,60206	12,97484	4,39651		— 3,67957	+ 0,41839
4	16	4,06778		— 0,65441	+ 0,08966
4,69897	22,08032	3,42308		+ 5,42591	— 0,55504
5	25	3,07664		+ 8,34559	— 0,90148
5,47712	29,99885	2,47857		+ 13,34444	— 1,49955
Medie	16,65441	3,97812	*Somme*	— 27,11597	+ 2,95606
				+ 27,11594	— 2,95607

(1) Un altro inconveniente, sotto l'aspetto formale, è nella scelta dei tre punti, la quale scelta è affidata al *prudente arbitrio* del calcolatore.

Il calcolo della costante c nell'equazione: $\log (y + c) = A - \alpha \log x$, si fa per tentativi nel modo seguente:

Si prendono tre valori di y, per esempio, 63.997, 11.689 e 301, corrispondenti ai valori di x: 1000, 10.000, 300.000 e si stabiliscono le equazioni:

$$* \log (63.997 + c) = A - \alpha \log 1000 = A - 3\alpha$$
$$\log (11.689 + c) = A - \alpha \log 10.000 = A - 4\alpha$$
$$\log (301 + c) = A - \alpha \log 300.000 = A - 5{,}47712\alpha.$$

Sottraendo la seconda equazione dalla prima e la terza dalla seconda, si avrà:

$$** \log (63.997 + c) - \log (11.689 + c) = \alpha$$
$$\log (11.689 + c) - \log (301 + c) = 1.47712\alpha.$$

Dividendo 2,95606 per 27,11597 si ottiene il quoziente — 0,109016 che è il valore di α. L'equazione cercata per questa via è quindi:

$$\log y - 3,97812 = -0,109016 (\log x^2 - 16,65441)$$

ossia:

$$\log y = 5,79371 - 0,109016 (\log x)^2.$$

Calcolando allo stesso modo per gli altri due paesi, le formole da confrontarsi risultano:

Per l'*Italia* . . . . . $\log y = 5,7937 - 0,1090 (\log x)^2$
Per la *Francia* . . . $\log y = 6,3403 - 0,1017 (\log x)^2$
Per l'*Inghilterra* . . $\log y = 5,7368 - 0,0769 (\log x)^2$

Le costanti 0,1090; 0,1017; 0,0769 sono precisamente quelle che caratterizzano la ripartizione della ricchezza nei tre paesi. In conformità di ciò che abbiamo detto sopra, quanto maggiore è α, coefficiente di

---

Ed ora dividendo la seconda di queste espressioni per la prima, otteniamo

$$\frac{\log (11.689 + c) - \log (301 + c)}{\log (63.997 + c) - \log (11.689 + c)} = 1,47712.$$

I tentativi cominciano qui. Diamo, per esempio, a c il valore 100. Si avrebbe:

$$\frac{\log 11.789 - \log 401}{\log 64.097 - \log 11.789} = 1,996\ldots$$

Il quoziente 1,996... è troppo elevato in confronto di quello di 1,477... al quale vogliamo arrivare. Allora proviamo per c il valore 500. Avremo:

$$\frac{\log 12.189 - \log 801}{\log 64.497 - \log 12.189} = 1,634\ldots$$

È ancora troppo. Tenteremo allora per c il valore 700; il risultato discende a 1,52... Siamo già vicini a quello voluto. Si provi il valore 800; si otterrà il quoziente 1,475...; il che vuol dire che ci siamo spinti un po' troppo innanzi. Il valore 795 è quello che meglio conviene.

Stabilito così il valore di $c = 795$, la equazione sopra segnata coll'asterisco doppio (**) ci dà tosto il valore di α

$$\log (63.997 + 795) - \log (11.689 + 795) = \alpha$$

donde:

$$\alpha = 0,71517.$$

E conosciuto così il valore di α, si determina quello di A coll'equazione segnata sopra coll'asterisco semplice (*):

$$\log (63.997 + 795) = A - 3 \times 0,71517$$

donde:

$$A = 6,95703.$$

Quindi l'equazione cercata della seriazione italiana si può scrivere:

$$\log (y + 795) = 6,95703 - 0,71517 \log x$$

e liberandola dai logaritmi:

$$y = \frac{A}{x^{0,71517}} - 795$$

dove con A si indica la costante il cui logaritmo è 6,95703.

$(\log x)^2$, tanto minore deve ritenersi la disuguaglianza nella ripartizione; e viceversa. L'Inghilterra ci appare dunque, come già appariva dal diagramma, un paese in cui il numero dei ricchi non decresce col crescere dei patrimoni così rapidamente come in Francia e ancor più in Italia. Il quasi parallelismo delle curve della Francia e dell'Italia ha la sua espressione numerica nella quasi parità dei coefficienti 0,1090 e 0,1017; e sta a dimostrare che questi due paesi, nonostante la grande differenza di ricchezza assoluta, si somigliano molto per ciò che concerne la legge della ripartizione di essa tra gli strati sociali.

È doveroso avvertire che le formole date rappresentano soltanto una prima approssimazione; l'elevazione a quadrato dei $\log x$ non produce una curvatura della linea interpolatrice altrettanto pronunciata, quanto quella della linea reale.

L'errore però avviene in egual senso per tutti tre i paesi. Converrebbe forse interpolare per tutti tre i paesi la funzione $(\log x)^{2,15}$, perchè l'approssimazione riuscisse per tutti tre più soddisfacente, ***pur rimanendo una sola la costante caratteristica vincolata alla variabile.*** Quest'ultima è una condizione necessaria della comparazione dei risultati, una condizione da realizzare anche con qualche scapito dell'approssimazione. La brevità, che ci siamo imposti, vieta d'entrare in ulteriori calcoli e discussioni; basti aver dato qui un saggio della estrema semplicità e della sintesi, cui si possono ridurre considerevoli masse di dati mercè facili procedimenti interpolatorii, preceduti da opportune rappresentazioni grafiche su scala logaritmica, mentre l'occhio scorrendo file di numeri nelle tabelle statistiche non avrebbe saputo cogliere gli aspetti loro veramente caratteristici.

§ **3.** Le curve del genere binomiale non sono che il limite, verso cui tende il poligono descritto coi termini di sviluppo del binomio $(p+q)^m$, quando m si fa diventare sempre più grande. Limitiamoci a dare una idea generale dell'argomento.

È noto che:

$$
\begin{aligned}
(p+q)^2 &= p^2 + 2pq + q^2 \\
(p+q)^3 &= p^3 + 3p^2q + 3pq^2 + q^3 \\
(p+q)^4 &= p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4 \\
(p+q)^5 &= p^5 + 5p^4q + 10p^3q^2 + 10p^2q^3 + 5pq^4 + q^5 \\
(p+q)^6 &= p^6 + 6p^5q + 15p^4q^2 + 20p^3q^3 + 15p^2q^4 + 6pq^5 + q^6, \text{ ecc.}
\end{aligned}
$$

La legge di formazione dei termini è semplicissima. Gli esponenti di p diminuiscono di una unità ad ogni successivo termine; quelli di q aumentano man mano di una unità. I coefficienti poi non sono altro che le combinazioni di m elementi presi uno ad uno, due a due, tre a tre, ecc., come meglio vedremo parlando del calcolo combinatorio.

Supponiamo $p = 1$, $q = 1$; e facciamo le sesta potenza del binomio. Sarà:

$$(1+1)^6 = 1 + 6 + 15 + 20 + 15 + 6 + 1 = \mathbf{64}.$$

Questa seriazione simmetrica, descritta graficamente, dà luogo al poligono A nella figura qui sotto.

Sia ora p triplo di q, e precisamente

$$p = \frac{3}{2}, \quad q = \frac{1}{2}$$

in modo da avere $p + q = 2$, come poc'anzi.

Elevando alla sesta potenza, si avrà:

$$\left(\frac{3}{2} + \frac{1}{2}\right)^6 = \left(\frac{3}{2}\right)^6 + 6\left(\frac{3}{2}\right)^5\left(\frac{1}{2}\right) + 15\left(\frac{3}{2}\right)^4\left(\frac{1}{2}\right)^2 +$$
$$+ 20\left(\frac{3}{2}\right)^3\left(\frac{1}{2}\right)^3 + 15\left(\frac{3}{2}\right)^2\left(\frac{1}{2}\right)^4 + 6\left(\frac{3}{2}\right)\left(\frac{1}{2}\right)^5 + \left(\frac{1}{2}\right)^6 =$$
$$= 11{,}39 + 22{,}78 + 18{,}98 + 8{,}44 + 2{,}11 + 0{,}28 + 0{,}01 = 64.$$

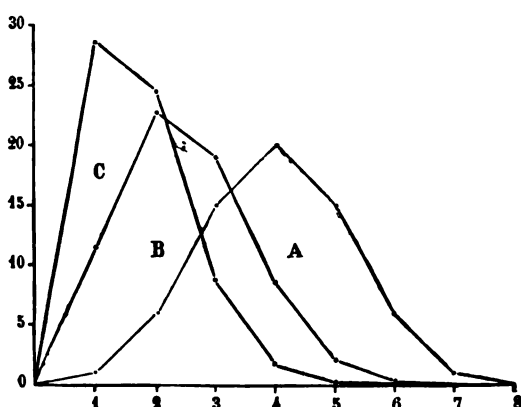
Questa seriazione, che evidentemente non è più simmetrica, dà luogo graficamente descritta al poligono B.

Supponiamo infine p sette volte più grande di q, e precisamente:

$$p = \frac{7}{4}, \quad q = \frac{1}{4},$$

sicchè si abbia ancora $p + q = 2$. La sesta potenza darà:

$$\left(\frac{7}{4} + \frac{1}{4}\right)^6 = \left(\frac{7}{4}\right)^6 + 6\left(\frac{7}{4}\right)^5\left(\frac{1}{4}\right) + 15\left(\frac{7}{4}\right)^4\left(\frac{1}{4}\right)^2 +$$
$$+ 20\left(\frac{7}{4}\right)^3\left(\frac{1}{4}\right)^3 + 15\left(\frac{7}{4}\right)^2\left(\frac{1}{4}\right)^4 + 6\left(\frac{7}{4}\right)\left(\frac{1}{4}\right)^5 + \left(\frac{1}{4}\right)^6 =$$
$$= 28{,}72 + 24{,}62 + 8{,}79 + 1{,}67 + 0{,}18 + 0{,}01 + 0{,}0\ldots = 64.$$

Tale seriazione, ancor più asimmetrica della precedente, è rappresentata dal poligono C.

Ordinate dei poligoni

A	B	C
1	11,39	28,72
6	22,78	24,62
15	18,98	8,79
20	8,44	1,67
15	2,11	0,18
6	0,28	0,01
1	0,01	0,00..

Si comprende che, se in luogo della sesta potenza, rappresentassimo lo sviluppo del binomio alla 10ª, 20ª o ad altra potenza superiore, adattando le scale in modo d'avere la stessa superficie di poligoni, il perimetro di questi si confonderebbe a poco a poco con una curva continua.

Il termine generale di sviluppo del binomio è

$$ {}^{x}C_{m}\, p^{m-x} \,.\, q^{x} $$

dove, secondo una notazione che noi preferiamo, ${}^{x}C_{m}$ indica il numero delle combinazioni di m elementi, presi zero a zero (ossia m ad m), uno ad uno, due a due, ecc., giusta il valore che si fa assumere alla variabile x. Ponendo, ad esempio, $m = 6$; $p = q = 1$ e dando alla x i successivi valori 0, 1, 2, 3... otteniamo precisamente i valori delle ordinate del poligono A. L'origine delle x è qui al principio della serie (1).

Orbene, quando una seriazione statistica, rappresentata in diagramma ordinario, prende aspetto d'un poligono binomiale, l'equazione della curva interpolatrice si determina facilmente prendendo i logaritmi dei numeri dati e interpolando una parabola di secondo grado:

$$ \log y = a + bx + cx^2. $$

Infatti, se si tratta proprio di una curva limite di un poligono binomiale, i logaritmi dei numeri osservati debbono avere costanti le differenze seconde, cioè costituire una serie parabolica di secondo grado (2).

---

(1) Se l'origine delle x si fissasse alla metà esatta della seriazione, il termine generale diverrebbe:

$$ y = {}^{\frac{m}{2}+x}C_{m} \,.\, p^{\frac{m}{2}-x} \,.\, q^{\frac{m}{2}+x} = {}^{\frac{m}{2}+x}C_{m} \,.\, \left(\frac{q}{p}\right)^{x} \,.\, (pq)^{\frac{m}{2}}. $$

(2) Prendiamo infatti i logaritmi della seconda equazione stabilita nella precedente nota. Sarà:

$$ \log y = \log \left({}^{\frac{m}{2}+x}C_{m}\right) + x \log \left(\frac{q}{p}\right) + \frac{m}{2} \log (pq). $$

Ma l'espressione $\left({}^{\frac{m}{2}+x}C_{m}\right)$ col crescere di m ha per limite $A e^{-h^2 x^2}$, dove A è l'ordinata massima della seriazione binomiale simmetrica ed e (base dei logaritmi naturali) $= 2.718...$ Sarà dunque

$$ \log \left({}^{\frac{m}{2}+x}C_{m}\right) = \log A - h^2 h^2 \log e; $$

e quindi

$$ \log y = \log A - h^2 x^2 \log e + x \log \left(\frac{q}{p}\right) + \frac{m}{2} \log (pq) $$

e facendo:

$$ \log A + \frac{m}{2} \log (pq) = a; \qquad \log \left(\frac{q}{p}\right) = b; \qquad - h^2 \log e = c, $$

si ottiene:

$$ \log y = a + bx + cx^2 $$

che è precisamente l'equazione di una parabola descritta coi logaritmi di y

Vogliasi, ad esempio, esprimere analiticamente la seriazione degli arruolati italiani per indice cefalico, data a pag. 111. Escludendo per semplicità di calcolo le classi a limiti indeterminati, cioè quella con indice superiore a 96 e quella con indice inferiore a 70, fisseremo l'origine delle x all'indice 82-84, cui corrisponde la classe centrale e di maggior frequenza. Indi, presi i logaritmi dei numeri dati, interpoleremo una parabola col metodo dei minimi quadrati o con altro metodo più spedito. Col primo metodo si arriva all'equazione:

$$\log y = 2{,}81699 - 0{,}01425\ \psi_{I} - 0{,}088034\ \psi_{II}$$

ed essendo

$$\psi_{I} = x; \qquad \psi_{II} = x^2 - \frac{n^2 - 1}{12}; \qquad n = 9,$$

avremo:

$$\log y = 3{,}40388 - 0{,}01425\ x - 0{,}088034\ x^2.$$

Confrontando ora i dati dell'osservazione con quelli teorici, ottenibili da cotesta formola, abbiamo:

OSSERVAZIONE			CALCOLO		Differenze tra il calcolo e l'osservazione	
x	y	log y	log y	y	Nei log y	Negli y
— 4	103	2,01284	2,05234	112,8	+ 0,03950	+ 9,8
— 3	463	2,66558	2,65433	451,2	— 0,01125	— 11,8
— 2	1.261	3,10072	3.08025	1203,0	— 0,02047	— 58,0
— 1	2.020	3,30535	3,33010	2138.5	+ 0,02475	+ 118,5
0	2.702	3,43169	3,40388	2534,4	— 0,02781	— 167.6
1	1.898	3,27830	3,30160	2002.6	+ 0,02330	+ 104,6
2	1.086	3,03583	3,02325	1055,0	— 0.01258	— 31,0
3	347	2,54033	2,56883	570,5	+ 0,02850	+ 23,5
4	96	1,98227	1.93834	86.8	— 0.04393	— 9,2
	Medie	2,81699	2,81699			

Se si tien conto della circostanza, che noi abbiamo operato su una serie greggia, affetta certo da non pochi errori imputabili alla nota simpatia dei misuratori per le cifre rotonde, gli scostamenti tra il calcolo e l'osservazione ci devono apparire piccoli, anzichenò. Il loro alternarsi di positivi e negativi è altro indizio della imprecisione dei rilievi originarii.

Il peso speciale che acquistano in calcolo i logaritmi delle classi estreme, grazie al metodo dei minimi quadrati, nuoce alquanto alla concordanza tra i numeri osservati e i calcolati; per controbilanciarne gli effetti si suggerisce di calcolare successivamente la formola su un numero crescente di termini, ad esempio, prima sui cinque termini centrali; poi su sette, infine su tutti nove, prendendo la media aritme-

tica delle costanti ottenute. Il calcolatore esercitato vede presto lo spediente più adatto al caso.

Uno spediente di tal genere apparirebbe anche più necessario, se si tenesse conto dei gruppi estremi a limiti indeterminati, coll'assegnarli a classi in qualche modo definite. Così i dodici individui dall'indice cefalico inferiore a 70 potrebbero attribuirsi tutti alla classe 67-69; ovvero quasi tutti a questa ed alcuni alla classe 64-66; e analogamente dicasi per gli individui dall'indice superiore a 96. Certo, cotesta assegnazione, si eseguisca ad occhio o con qualche artificio matematico, involge un po' d'arbitrio; ma offre il vantaggio di introdurre in calcolo tutti gli elementi della seriazione osservata.

È ovvio che, se le classi fornite dalle statistiche sono inegualmente estese, occorrerà un lavoro preparatorio di interpolazioni parziali, allo intento che la seriazione proceda per intervalli eguali dei valori della variabile.

La riduzione delle curve binomiali a parabole ordinarie, descritte coi logaritmi dei dati osservati, si raccomanda per la sua semplicità e per la facilità colla quale si presta a risolvere quesiti di questo genere: come si ripartiscono gli individui nelle sottoclassi, in cui possiamo immaginare frazionata una certa classe: qual'è il valore della variabile, all'uno o all'altro degli estremi della seriazione, per cui si ha un solo individuo del gruppo teorico (1): qual'è il valore della variabile, in corrispondenza del quale si ha il massimo valore della funzione (2); e

---

(1) Se ad un certo valore della x deve corrispondere un solo individuo del gruppo, sarà: $y = 1$; $\log y = 0$. E secondo che si considera l'estremo a destra o quello a sinistra dell'ordinata elevata sull'origine, cioè secondo che si fa x positivo o negativo, l'equazione (prendendo, ad esempio, quello che si riferisce agli arruolati distinti secondo l'indice cefalico) sarà:

$$3{,}40388 - 0{,}01425\,x - 0{,}088034\,x^2 = 0$$

oppure:

$$2{,}40388 + 0{,}01425\,x - 0{,}088034\,x^2 = 0$$

equazioni di secondo grado, di cui è facile la soluzione.

(2) Questo problema, risolubile algebricamente per tentativi con metodo analogo a quello indicato a pag. 163 e seguenti, riesce ancor più facile a chi abbia le prime nozioni del calcolo infinitesimale. Per rendere

$$3{,}40388 - 0{,}01425\,x - 0{,}088034\,x^2 = maximum$$

si faccia la derivata:

$$-0{,}01425 - 2 \times 0{,}088034\,x = 0$$

donde:

$$x = -0{,}08.$$

Ora, siccome nell'esempio nostro ogni unità d'intervallo di valori della variabile si era fatta corrispondere a 3 centesimi d'indice cefalico, il valore di x, espresso in centesimi d'indice, sarà $-0{,}24$; e supponendo $= 83$ l'indice mediano della classe 82-84, si avrà: $83 - 0{,}24 = 82{,}76$. Quindi, secondo la formola, all'indice cefalico di 82,76 % si avrebbe l'ordinata massima della curva.

così via. Infine agevola abbastanza i confronti tra seriazioni della stessa specie, quantunque le costanti vincolate alla variabile siano due e non una. All'uopo non sarà inutile ricordare che il coefficiente di x indica il grado di asimmetria della seriazione, quando l'origine sia riferita al punto, cui corrisponde il *maximum* di frequenza; il coefficiente di x^2 indica invece la rapidità colla quale decrescerebbero le successive ordinate, a destra e a sinistra dell'ordinata massima, se non intervenisse quell'elemento perturbatore che, rappresentato dal termine in x, produce l'asimmetria della curva seriale.

E) *Dell'interpolazione nel caso di osservazioni mediate.*

Sommario: § 1. Metodo per la determinazione di *equazioni normali* nel caso di equazioni empiricamente stabilite in numero superiore alle incognite.

§ **1.** In fisica, in geodesia, ecc., avviene spesso che le osservazioni da cui si possono ricavare le costanti di una equazione siano in numero maggiore delle incognite, sì che il problema riesce più che determinato.

Allora è necessario combinare le diverse equazioni per guisa da formarne delle nuove in numero eguale a quello delle costanti da determinarsi, alla condizione solita che la somma dei quadrati degli errori delle singole osservazioni sia un minimo. Noi sappiamo, ad esempio, dalla meccanica che la lunghezza l del pendolo a secondi è:

$$l = A + B \cos^2 \varphi,$$

ove φ è la latitudine del luogo ed A e B sono due costanti da determinarsi. Ora se la misura della lunghezza del pendolo e il rilievo della latitudine potessero farsi in modo matematicamente esatto e se si fosse assolutamente sicuri della eliminazione di circostanze perturbatrici, basterebbero due osservazioni a stabilire il preciso valore delle due costanti.

Ma poichè in pratica errori di varia natura si verificano sempre, convien moltiplicare le osservazioni, salvo poi a combinare le equazioni corrispondenti in guisa da ridurle ancora, nel caso concreto, a due, cioè a tante, quante sono le incognite.

Il metodo è semplicissimo: consiste nel moltiplicare ambo i membri di ciascuna equazione prima per il coefficiente di una delle incognite, poi per il coefficiente dell'altra incognita e fare le somme rispettive, che forniranno le cosidette *equazioni normali.*

Supponiamo, per venire ad un caso pratico, che una certa misura o determinazione debba soddisfare all'equazione:

$$a = x + n\, y.$$

E si siano ottenuti in cinque osservazioni, eseguite in circostanze diverse, i seguenti valori:

$$0 = x + 1{,}50\,y$$
$$8{,}95 = x + 6{,}75\,y$$
$$18{,}50 = x + 11{,}00\,y$$
$$27{,}50 = x + 15{,}50\,y$$
$$35{,}50 = x + 20{,}50\,y$$

Qui il coefficiente di x è sempre l'unità; perciò, moltiplicando per 1 ambo i membri delle singole equazioni, si avrà ancora la serie data. Invece i coefficienti di y essendo 1,50, 6,75, 11, ecc., prendendoli come moltiplicatori delle equazioni in cui figurano, otterremo una serie modificata:

	Moltiplicando per i coefficienti di x	Moltiplicando per i coefficienti di y
	$0 = x + 1{,}50\,y$	$0 = 1{,}50\,x + 2{,}25\,y$
	$8{,}95 = x + 6{,}75\,y$	$60{,}41 = 6{,}75\,x + 45{,}56\,y$
	$18{,}50 = x + 11{,}00\,y$	$203{,}50 = 11{,}00\,x + 121{,}00\,y$
	$27{,}50 = x + 15{,}50\,y$	$426{,}25 = 15{,}50\,x + 240{,}25\,y$
	$35{,}50 = x + 20{,}50\,y$	$727{,}75 = 20{,}50\,x + 420{,}25\,y$
Somme	$90{,}45 = 5\,x + 55{,}25\,y$	$1417{,}91 = 55{,}25\,x + 829{,}31\,y$

Quindi le equazioni normali saranno:

$$5\,x + 55{,}25\,y = 90{,}45$$
$$55{,}25\,x + 829{,}31\,y = 1417{,}91,$$

da cui si ricava: $x = -3{,}04$; $y = 1{,}91$ (1).

I quali valori soddisfano alla condizione dei minimi quadrati.

Un risultato quasi eguale si sarebbe ottenuto con un metodo più lungo, combinando due a due le cinque equazioni date (il che avrebbe fornito dieci diverse coppie di equazioni) e facendo la media aritmetica dei valori rispettivamente trovati per le due incognite.

---

(1) Abbiamo preso questo esempio dal FERRERO, *Esposizione del metodo dei minimi quadrati*, pag. 121 e seguenti. E a detta opera rimandiamo per la illustrazione matematica del procedimento indicato. L'esempio riguarda la determinazione del valore delle parti di una livella a bolla d'aria. Nella equazione $a = x + n\,y$, x è l'inclinazione iniziale del livello; y il valore di una parte della divisione del livello; n, n', n''... le letture del livello, cioè le medie delle posizioni estreme della bolla; a, a', a''... le letture corrispondenti fatte sul circolo verticale.

F) *Teoria delle correlazioni.*

Sommario: § 1. Concetto di *correlazione*. — § 2. Metodo per il calcolo della correlazione *semplice*. — § 3. Id. per il calcolo della correlazione *doppia*. — § 4. Rappresentazione grafica della « legge di regressione » del Galton. — § 5. Rapporti tra serie.

§ 1. Quando un fenomeno collettivo suol variare in una certa direzione e misura col variare di un altro fenomeno in una cert'altra direzione e misura, diciamo esistere tra i due una *correlazione*, nel senso che l'uno è causa o effetto dell'altro, oppure che entrambi dipendono da una causa comune che li fa variare concordemente. Però casi di correlazione perfetta non si dànno mai nell'ordine dei fatti collettivi; il che si spiega per la grande molteplicità di cause che entrano in giuoco nelle serie statistiche.

Il problema delle correlazioni è essenzialmente vincolato a quello dell'interpolazione, di cui pertanto riflette gli aspetti arbitrarii. « Correlazione » è infatti il rapporto che si stabilisce tra le parti variabili e concomitanti di due o più serie, prescindendo dalla parte presunta costante e indipendente in ciascuna serie. Ma come distinguere le parti variabili, per metterle a reciproco confronto, senza peccare d'arbitrio, dal momento che i processi d'interpolazione presentano, come abbiamo veduto, degli aspetti arbitrarii? Evidentemente si avranno tanti valori diversi per la correlazione, quanti i modi di interpolare; perocchè a seconda che il calcolatore riterrà rappresentata la parte regolare e autonoma della serie da una retta o da una parabola di secondo o di terzo grado, ecc., le oscillazioni residue, cioè le parti variabili sulle quali si esercita il calcolo di correlazione, risulteranno diverse. Peraltro, eseguito il maggior numero possibile di tentativi, la scelta non potrà non cadere su quel metodo, che realizza un *maximum* di concordanza tra le parti variabili dei fenomeni confrontati; ed è questo l'unico limite che forse esiste all'arbitrarietà del procedimento.

Esempi di correlazioni si trovano in ogni campo di osservazione statistica. Havvi correlazione tra prezzi e consumo delle merci, tra importazioni ed esportazioni di un paese, tra matrimoni e nascite legittime, fra la densità della popolazione e l'altitudine dei luoghi abitati, fra la statura e la frequenza delle pulsazioni, fra il colore degli occhi e quello dei capelli.

La correlazione può essere di varii gradi: semplice, doppia, tripla... È *semplice* quella che si stabilisce fra le parti variabili di *due* sole serie, riferibili a due fenomeni, la cui concomitanza di variazioni si abbia ragion di credere non accidentale; *doppia* quella che si stabilisce fra *tre* serie di fenomeni, di cui si presume la reciprocità di influenza; e così via. Così la natalità legittima, considerata specialmente nelle sue

variazioni mensili sembra riflettere ad un tempo le oscillazioni della nuzialità e quelle della mortalità di nove o dieci mesi addietro; ecco un caso di probabile correlazione doppia.

§ 2. *Metodo elementare per il calcolo della correlazione semplice.* — Siano i fenomeni A e B rappresentati nelle due serie:

A	35	31	32	31	40	35
B	210	166	182	174	241	185

Quando A diminuisce, diminuisce pure B; quando il primo aumenta, aumenta il secondo. Vogliamo precisare di quante unità aumenta o diminuisce A coll'aumentare o diminuire di un'unità di B.

Il procedimento più elementare e comune consiste nel prendere la media aritmetica delle due serie e calcolare gli scostamenti positivi o negativi di ciascun termine dalla media della rispettiva serie. Indi sommati gli scostamenti positivi (o negativi) della serie B e i *corrispondenti* scostamenti della serie A, dividiamo la seconda somma per la prima. Il quoziente trovato è il coefficiente di correlazione semplice.

Nel caso in esame dunque avremo:

	A		B	δA	δB
	35		210	+ 1	+ 17
	31		166	— 3	— 27
	32		182	— 2	— 11
	31		174	— 3	— 19
	40		241	+ 6	+ 48
	35		185	+ 1	— 8
Media	34	*Media*	193		

Somma dei δB positivi	Somma dei corrispondenti δA	Somma dei δB negativi	Somma dei corrispondenti δA
+ 17	+ 1	— 27	— 3
+ 48	+ 6	— 11	— 2
—	—	— 19	— 3
—	—	— 8	+ 1
+ 65	+ 7	— 65	— 7

$$7 : 65 = x : 1,$$

donde

$$x = 0{,}108.$$

Ciò vuol dire che per ogni unità in più o in meno della media di B si ha una variazione di 0,108 unità in più o in meno della media di A. Sicchè possiamo esprimere A in funzione della propria media e delle variazioni di B, così:

$$A = 34 + 0{,}108\ \delta\ B.$$

Supponendo, ad esempio, che B nell'anno successivo all'ultimo della serie data diventasse = 200, ossia superasse di sette unità la media del periodo già osservato, A dovrebbe essere

$$= 34 + 0{,}108 \times 7 = 34{,}76.$$

Se B scendesse a 153, ossia scendesse di 40 punti sotto la media già osservata, dovremmo avere:

$$A = 34 + 0{,}108 \times (-40) = 29{,}64.$$

Da ciò si può arguire la competenza del calcolo delle correlazioni nei casi in cui si tratti di colmare lacune nelle serie. Avviene spesso che serie statistiche si trovino interrotte, perchè in un dato anno o periodo non si vollero eseguire i necessari rilievi; ma le lacune si possono colmare, in via di approssimazione, se per quell'anno o periodo esistono i dati concernenti altro fenomeno, di cui si conosca il grado medio di correlazione col primo.

Nell'esempio sopra esposto è chiaro che se la correlazione tra A e B fosse perfetta, la formola: $A = 34 + 0{,}108\ \delta B$ dovrebbe riprodurre esattamente i valori di A forniti dall'osservazione. La correlazione invece appare imperfetta, esistendo tra i valori osservati di A e i valori teorici, calcolati coll'anzidetta formola, piccole differenze, le quali lasciano sospettare essere le variazioni di A dipendenti non solo dalle variazioni di B, ma ancora da altre cause conoscibili o non conoscibili. Infatti, applicando la formola, otteniamo:

Valori teorici di A	Valori osservati	Differenze tra il calcolo e l'osservazione
34 + 0,108 × 17 = 35,84	35	— 0,84
34 + 0,108 × (— 27) = 31,08	31	— 0,08
34 + 0,108 × (— 11) = 32,81	32	— 0,81
34 + 0,108 × (— 19) = 31,95	31	— 0,95
34 + 0,108 × 48 = 39,18	40	+ 0,82
34 + 0,108 × (— 8) = 33,14	35	+ 1,86

Dal momento che esistono differenze, or negative, or positive, tra l'osservazione e il calcolo, segno è che vi sono nella serie di A degli eccessi o dei difetti di oscillazione non spiegabili colla sola influenza di B; si è condotti perciò a sospettare la presenza di qualche altra causa.

§ 3. *Calcolo della correlazione doppia.* — Immaginiamo dunque di sapere che un terzo fenomeno C è probabilmente collegato pur esso con A per via di un nesso causale. Le variazioni di C saranno allora in un certo grado concomitanti con quelle di A; e poichè A presenta una certa concomitanza con B, così anche tra C e B potrà esserci analogia di variazioni. Siano, ad esempio, per B e C le serie:

	Serie B	Serie C
	210	98
	166	97
	182	99
	174	92
	241	106
	185	108
Media	193	100

Facendo, col metodo ora esposto, il calcolo di correlazione fra C e B, si arriva alla formola:

$$C = 100 + 0,061\ \delta B.$$

L'applicazione di questa formola dà per C dei valori teorici che, confrontati con quelli forniti dall'osservazione, dimostrano esistere anche nella serie C eccessi e difetti di oscillazione non spiegabili colla sola influenza di B. Si avrebbe infatti:

Valori teorici di C	Valori osservati di C	Differenze
101,04	98	— 3,04
98,35	97	— 1,35
99,33	99	— 0,33
98,84	92	— 6,84
102,93	106	+ 3,07
99,51	108	+ 8,49

Eliminando dunque la parte d'influenza che B spiega su C, resterà a vedere se non ci sia per avventura una concomitanza fra le residue oscillazioni di C, le quali non si possono spiegare coll'influenza di B e le residue oscillazioni di A, che pure, come abbiam visto nel paragrafo precedente, non sono spiegabili colla influenza di B. Il confronto si stabilirà dunque nel seguente modo:

Residue oscillazioni di A non spiegabili coll'influenza di B	Residue oscillazioni di C non spiegabili coll'influenza di B
— 0,84	— 3,04
— 0,08	— 1,35
— 0,81	— 0,33
— 0,95	— 6,84
+ 0,82	+ 3,07
+ 1,86	+ 8,49

Il parallelismo è evidente. Sommando i termini negativi (o positivi) della seconda serie e i corrispondenti termini della prima e dividendo questa somma per quella si ha: $\frac{2,68}{11,56} = 0,232$. Il che vuol dire che per ogni unità di oscillazione residua di C, abbiamo 0,232 unità di residua oscillazione in A. Chiamando con rC le oscillazioni residue di C, si stabilirà l'equazione:

$$A = 34 + 0,108\ \delta B + 0,232\ rC.$$

Applicando la qual formola si ottengono i seguenti valori teorici per A in funzione di δB e di rC.

Valori teorici di A in funzione di δB e rC	Valori osservati di A	Differenze
35,84 + 0,232 × (— 3,04) = 35,13	35	— 0,13
31,08 + 0,232 × (— 1,35) = 30,77	31	+ 0,23
32,81 + 0,232 × (— 0,33) = 32,73	32	— 0,73
31,95 + 0,232 × (— 6,84) = 30,36	31	+ 0,64
39,18 + 0,232 × 3,07 = 39,89	40	+ 0,11
33,14 + 0,232 × 8,49 = 35,11	35	— 0,11

Si vede ora che le differenze tra la serie calcolata di A e la serie osservata sono molto minori di quanto si era tenuto calcolo della sola influenza del fenomeno B.

Essendo $rC = C - (100 + 0{,}061\, \delta B)$ la formola testè ottenuta può convertirsi in quest'altra:

$$A = 34 + 0{,}108\, \delta B + 0{,}232\, (C - 100 - 0{,}061\, \delta B)$$

e poichè: $C - 100 = \delta C$, si avrà:

$$A = 34 + 0{,}108\, \delta B + 0{,}232\, (\delta C - 0{,}061\, \delta B)$$

che può ancora semplificarsi così:

$$A = 34 + 0{,}094\, \delta B + 0{,}232\, \delta C.$$

Questo il procedimento che ci interessava indicare per il calcolo della *correlazione doppia.*

Le oscillazioni residue di A, cioè gli eccessi e i difetti di oscillazione che non son spiegabili nemmeno coll'influenza combinata di B e di C, potrebbero dipendere da un quarto fattore D. A voler tener conto anche di questo, si dovrebbero confrontare le oscillazioni residue di A, testè calcolate, con quelle che presenta D *dopo di essere stato posto alla sua volta in doppia correlazione* con B e con C; e indi procedere in maniera analoga a quella dianzi segnata. Si avrebbe così una formola di *correlazione tripla.*

Gli è evidente che quanto più lunghe sono le serie e quanto maggiore il numero dei fenomeni, di cui si misura l'azione combinata sul fenomeno dato, tanto più probabile sarà la precisione di questo, fondata sugli elementi conosciuti di quelli. Finchè infatti si considerano correlazioni semplici, A ci potrà sembrare tratto tratto come svincolato dalla influenza di B, causa parziale tra le molte che certo concorrono a regolare il suo movimento; ma raro avverrà che esso si svincoli ad un tempo da B, C, D, ecc. Così il consumo del caffè varia di solito in una certa ragione inversa del suo prezzo; se talvolta sembra sfuggire a questa disciplina, gli è perchè il consumo di quella derrata dipende in parte anche dai prezzi dello zucchero e della cicoria, beni d'uso complementare, in parte dai prezzi della birra, dell'alcool, ecc., beni d'uso concomitante o concorrente; e infine dai prezzi di altre merci ancora, caratteristiche del tenor di vita delle classi, che sono del caffè le maggiori consumatrici. Sarebbe desiderabile appunto possedere per le derrate più importanti le formole empiriche che ne presentassero il consumo come funzione dei prezzi loro proprii e dei prezzi dei beni d'uso complementare, concomitante o concorrente; formole siffatte entrando come elementi di calcolo nella soluzione di problemi di economia applicata, costituirebbero la migliore utilizzazione immaginabile dell'immenso materiale statistico intorno ai prezzi e ai consumi, di cui oggi non si sfrutta la centesima parte.

Allorchè i fenomeni considerati hanno carattere, non statico, ma dinamico, convien prendere le differenze tra ciascun termine e il successivo, diminuirle della propria media aritmetica (1) e indi procedere al calcolo di correlazione nel modo indicato. In questa guisa il Pareto ha calcolato per i 40 anni dal 1855 al 1895 la correlazione doppia tra il numero dei matrimoni, il valore delle esportazioni e la produzione del carbon fossile in Inghilterra. La sua formola, semplificata, sarebbe:

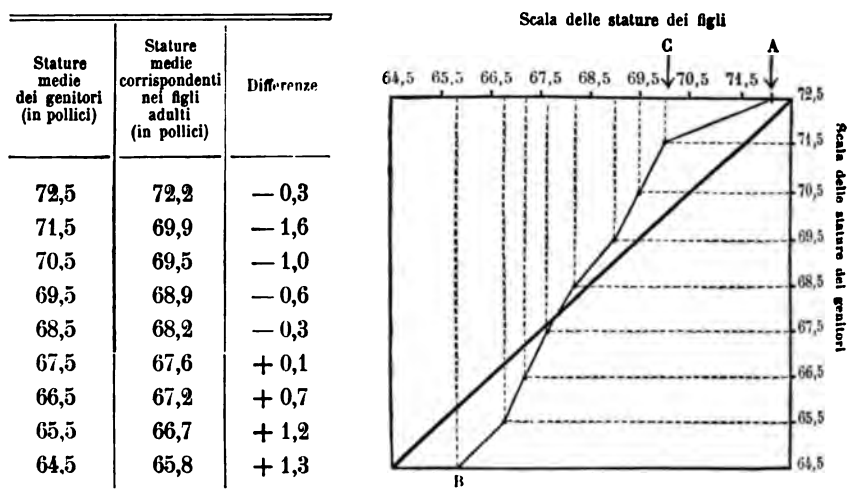
$$\delta v = 0{,}2499\ \delta u + 0{,}05523\ \delta t - 0{,}6824$$

dove v indica il numero dei matrimoni in migliaia; u il valore delle esportazioni in migliaia di sterline; t la quantità del carbon fossile prodotto, in centinaia di migliaia di tonnellate: δv, δu, δt le differenze da un anno all'altro (2).

Ciò che spesso nasconde in tutto o in parte la correlazione è l'intervallo di tempo che può correre tra la causa e l'effetto, tra la variazione del fenomeno che spiega un'influenza e la variazione del fenomeno che la subisce. È quindi necessario, a stabilire il vero sincronismo delle serie, un lavoro preparatorio attuato per tentativi razionali e confortato dalle nozioni che già si posseggono nella soggetta materia per via deduttiva o per osservazione anche non statistica.

§ 4. Il Galton ha suggerito un metodo assai semplice per rappresentare la correlazione in caso di seriazioni.

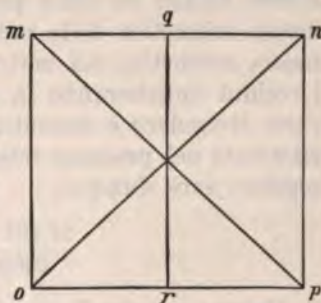
Figuriamoci di dover descrivere le serie correlative delle stature medie dei genitori e di quelle dei rispettivi figli adulti.

Stature medie dei genitori (in pollici)	Stature medie corrispondenti nei figli adulti (in pollici)	Differenze
72,5	72,2	— 0,3
71,5	69,9	— 1,6
70,5	69,5	— 1,0
69,5	68,9	— 0,6
68,5	68,2	— 0,3
67,5	67,6	+ 0,1
66,5	67,2	+ 0,7
65,5	66,7	+ 1,2
64,5	65,8	+ 1,3

(1) Vale a dire, facendo la somma algebrica di tali differenze e dividendo il risultato pel numero di esse; o, ciò che torna lo stesso, dividendo per questo numero la differenza fra il primo e l'ultimo termine della serie data.

(2) V. *Quelques exemples d'application*, ecc., già citato a pag. 179.

Per rendere paragonabili tra loro i dati concernenti famiglie diversamente assortite per sesso dei componenti, il Galton uniformò in certa guisa le stature delle femmine a quelle dei maschi, moltiplicandole per il coefficiente 1,08, essendo appunto la statura della donna rispetto a quella dell'uomo nel rapporto di 1:1,08. Sui dati così trasformati il Galton calcolò le medie riportate alla pagina precedente.

Evidentemente, se da genitori aventi la statura media di 72,5 derivassero figli aventi (nell'età adulta) la stessa statura media, l'origine della linea dei figli coinciderebbe coll'estremo della diagonale. Essendo invece la statura media dei figli alquanto inferiore e precisamente di 72,2, l'origine cadrà nel punto, verso cui è diretta la freccia A. Del pari, se a genitori di 71,5 corrispondessero figli di 71,5, bisognerebbe abbassare la perpendicolare dal punto segnato 71,5 nella scala delle stature dei figli, fino all'incontro colla orizzontale partente dal punto segnato 71,5 nella scala delle stature dei genitori; e l'incontro avverrebbe sulla diagonale. Invece, siccome, secondo il prospetto numerico del Galton, da genitori di 71,5 derivano figli di 69,9 pollici soltanto (beninteso *in media*), così la perpendicolare andrà abbassata dal punto, verso cui si dirige la freccia C, fino all'incontro colla orizzontale partente dal punto segnato 71,5 nella scala delle stature dei genitori. E così via. Si ottiene insomma la spezzata A B, la quale per un certo tratto passa sopra, poi passa sotto la diagonale. Tale spezzata (per cui si potrebbe interpolare facilmente una retta) descrive le stature medie dei figli per rapporto a quelle dei corrispondenti genitori. Ciascuno intende che se la correlazione fosse perfetta, cioè se da genitori di una qualsiasi statura derivassero figli del tutto identici a loro, la linea d'incontro coinciderebbe colla diagonale $o\,n$ del quadrato qui tracciato. Al contrario, se dai genitori più alti nascessero i figli più bassi, dai genitori più piccoli di statura i figli più alti, e dai medii i figli medii, la linea di incontro sarebbe rappresentata dalla diagonale opposta, $m\,p$. Allora la correlazione sarebbe ancora perfetta, ma di segno contrario; cioè se chiamiamo *positiva* la prima, dovremmo dire *negativa* la seconda. Infine, se correlazione non esistesse affatto e cioè da genitori di statura superiore alla media provenissero indifferentemente figli alti o bassi e da genitori inferiori alla media pure indifferentemente figli bassi o alti, la linea d'incontro delle stature dei genitori e dei figli sarebbe la $q\,r$.

Nel caso concreto la correlazione risulta positiva, ma non perfetta. Il suo valore ρ grado ha per espressione la tangente dell'angolo che la retta interpolata fra i punti della spezzata A B farebbe colla diagonale.

La legge galtoniana (detta *legge di regressione*) qui esposta si può formolare così: « Da genitori che, per statura, si discostano d'una certa quantità x, in più o in meno, dalla media normale della popolazione, provengono figli, che si discostano solo di $^2/_3$ di x da detta media ». La legge sussiste anche se da gruppi scelti di figli si risale ai genitori o da soggetti scelti nelle figliuolanze si passa ai loro fratelli (1).

§ 5. Affine all'argomento delle *correlazioni* è quello dei *rapporti tra serie*. Come si istituiscono rapporti tra due quantità osservate indipendentemente dal fatto che esista o non esista correlazione tra loro, così si possono istituire rapporti tra due serie, o, meglio, tra le espressioni analitiche in cui esse serie sono sintetizzate. Così, date le equazioni delle parabole interpolatrici della natalità legittima e della nuzialità in Italia dal 1872 al 1900 avremmo:

$$\frac{\text{Natalità legittima}}{\text{Matrimoni}} = \frac{1{,}021.964 + 3798{,}52\,x - 318{,}3\,x^2}{228.450 + 703{,}42\,x - 58{,}0\,x^2}.$$

Di qui si ricava come ha variato nel corso del periodo in esame la *fecondità totale* media per matrimonio. Basta all'uopo attribuire ad x i successivi valori 1, 2, 3 ... — 1, — 2, — 3 ..., l'origine essendo fissata all'anno centrale del periodo, cioè a 1886.

Espressi in questa forma, i rapporti tra serie permettono di determinare ancora la correlazione tra le parti variabili regolari delle serie stesse. Infatti se dalla prima delle due equazioni date togliamo la media aritmetica della natalità legittima (999.683) e dalla seconda la media aritmetica dei matrimoni osservati nel ventinovennio (224.390), i residui indicheranno la parte variabile regolare delle due serie (la parte irregolare e dipendente da cause accidentali presumendosi già eliminata nel processo interpolatorio). Il rapporto tra le parti variabili regolari sarà dunque:

$$\frac{22.281 + 3798{,}52\,x - 318{,}3\,x^2}{4060 + 703{,}42\,x - 58{,}0\,x^2}.$$

Orbene, mentre il quoziente di 999.683 per 224.390 = 4,46 esprime la fecondità totale media per matrimonio nell'insieme dei matrimoni, senz'altra distinzione, è facile vedere che nelle equazioni residue le costanti, divise ciascuna per la sua corrispondente, dànno quozienti più elevati, che oscillano fra 5,4 e 5,9. Infatti 22.281 : 4060 = 5,49; 3798,52 : 703,42 = 5,40 e infine 318,3 : 58,0 = 5,49. Ci sembra pertanto di dover concludere che i matrimoni in più o in meno della media generale, che si sono verificati nel corso del periodo 1872-1900 in Italia,

(1) Abbiamo cercato di volgarizzare nel modo più semplice possibile questa legge del Galton nei nostri *Principii di Demografia*, pag. 80-87.

aggiunsero o sottrassero alla media generale della natalità legittima il risultato di una fecondità più alta dell'ordinario; la qual cosa è pienamente conforme al fatto, che dove la nuzialità è in aumento, come riflesso del crescere della popolazione, gli sposi giovani hanno nell'aumento stesso una parte preponderante; mentre la loro parte diminuisce di fronte a quella degli sposi in età più matura dove la nuzialità o la popolazione sono in regresso (1).

---

(1) Avendo noi fissata l'origine delle x all'anno 1886, è ovvio che muovendo a ritroso da questa data per risalire al 1872, ritroveremo una popolazione e una nuzialità decrescenti anno per anno. Quindi qui si tratta di regresso apparente, il quale però non toglie che sussista la conclusione affermata nel testo. Le due parabole interpolatrici segnalerebbero il punto *maximum* nella natalità e nuzialità italiana fra il 1892 e il 1893, e da questa data comincierebbe il regresso reale.

CAPO QUARTO

Elementi di calcolo combinatorio e di calcolo delle probabilità.

TITOLO I.

ELEMENTI DI CALCOLO COMBINATORIO.

Sommario: § 1. Considerazioni generali. — § 2. *Disposizioni* di elementi di una stessa serie. — § 3. *Permutazioni.* — § 4. *Combinazioni.* — § 5. Combinazioni di elementi di serie diverse e indipendenti.

§ **1.** Per bene intendere la teoria delle probabilità e procedere poi alle necessarie applicazioni ai fatti statistici, occorrono alcune nozioni di calcolo combinatorio.

Accade continuamente di constatare che gli elementi di un aggregato, le unità di un gruppo, i termini di una serie si trovano in fatto disposti o combinati giusta un ordine definito, che è uno fra i tanti svariati ordini, secondo cui potremmo immaginare attuata la forma estrinseca della serie, del gruppo.

Osservo il cielo per tre giorni di seguito e tutti tre i giorni risultano sereni. Ecco un ordine determinato; eppure la natura ci dà esempi di seriazioni e combinazioni diverse, e quand'anche non ce li desse, potremmo sempre concepirli come possibili; così, oltre il caso dei tre giorni sereni, quello di tre coperti o di due sereni e uno coperto, in tre diversi modi di successione (SSC, SCS, CSS) o di due coperti e uno sereno, pure in tre diversi modi di successione (CCS, CSC, SCC).

Compongo con quattro lettere o suoni la parola «Roma». È una successione in serie ben definita e di significato ben noto. Ma potrei disporre quei quattro elementi in altre 23 maniere, formando altrettante parole, di cui alcune hanno pure un significato nella nostra lingua (*amor*, *ramo*, *orma*, ecc.) e le rimanenti non ne hanno punto.

Da un parto gemello possiamo attenderci due maschi o due femmine, o un maschio (come primo) e poi una femmina, o una femmina seguita da un maschio. Il caso in esame presenterà una delle quattro combinazioni possibili.

Gli esempi certo non difettano. Oggi la stessa chimica pura per dar ragione di alcune dissomiglianze fra corpi, i quali all'analisi si rivelano identici per composizione atomica, ricorre all'ipotesi che gli atomi in tali corpi siano diversamente disposti, l'uno come l'immagine

dell'altro veduta in uno specchio. Così il Pasteur spiegò le differenze di due acidi tartarici di egual composizione e tratti dalla stessa uva, segnalati dal Mitscherlich. Nella famiglia degli zuccheri sono frequenti queste forme *enantiomorfe*, che altrimenti si potrebbero dire forme « siamesi ».

Se la natura ci presentasse sempre combinazioni sotto tutte le forme assegnabili, indifferentemente, non sarebbe più il caso di parlare di « leggi esatte o empiriche ». Di queste leggi si parla, o perchè nei limiti della nostra esperienza la realtà in un dato àmbito di cose è rappresentata da certe combinazioni definite, ad esclusione di tutte le altre assegnabili a calcolo, o perchè, quando tutte le forme combinatorie possibili trovansi realizzate, lo sono con una frequenza di casi diversa, riducibile essa stessa a legge.

§ 2. Chiamansi *disposizioni* di m elementi, presi due a due, tre a tre... n ad n, i gruppi che si possono formare riunendo questi elementi due a due, tre a tre.... n ad n, a condizione di considerare come diversi l'un dall'altro i gruppi formati con elementi identici, ma diversamente disposti.

Siano m elementi (persone o cose, numeri o simboli), che designeremo colle lettere $a, b, c, d, \ldots$ Vogliamo sapere quante disposizioni si possono fare con questi m elementi presi 2 a 2.

La risposta è facile. Scelgo il primo elemento a e lo colloco volta per volta a sinistra di ciascuno dei restanti $m-1$ termini della serie, formando così $m-1$ gruppi binarii comincianti per a (ab, ac, ad, ecc.). La stessa cosa faccio con b, accompagnandolo con ciascuno degli altri $m-1$ elementi; otterrò dunque $m-1$ nuovi gruppi binarii comincianti per b (ba, bc, bd, ecc.). E così di seguito opero con c, con d, ecc., fino ad esaurire la serie. È chiaro pertanto che, essendo m gli elementi, e ciascuno di essi dando luogo, in combinazione con ogni altro, ad $m-1$ gruppi binarii, il totale di questi gruppi sarà $m\,(m-1)$.

Indicando con ${}^2D_m$ le disposizioni di m elementi presi 2 a 2, si avrà (1):

$${}^2D_m = m\,(m-1).$$

Per formare adesso le disposizioni di m elementi, presi 3 a 3, si scelga un qualunque gruppo binario, per esempio, ab, e gli si unisca volta per volta ciascuno degli $m-2$ elementi che rimangono. Si otterranno così $m-2$ gruppi ternarii comincianti per ab. Ma ciò che si è fatto con ab, può ripetersi per tutti gli $m\,(m-1)$ gruppi binarii, ognuno

(1) Esempi: 7 colori si possono disporre, due a due, in 42 maniere:

$$7 \cdot (7-1) = 42.$$

In 1560 maniere si possono disporre, due a due, le 40 carte di un mazzo:

$$40\,(40-1) = 1560.$$

dei quali concorrerà a formare $m - 2$ gruppi di 3 elementi o terne; queste terne saranno dunque in totale $m(m-1)(m-2]$. Ossia (1):

$$^3D_m = m(m-1)(m-2).$$

Analogamente si proverebbe che le disposizioni di m elementi, presi quattro a quattro, sono:

$$^4D_m = m(m-1)(m-2)(m-3).$$

E generalizzando, le disposizioni di m elementi presi n ad n sono date dalla formola:

$$^nD_m = m(m-1)(m-2)\ldots(m-n+1).$$

Esempio: 10 individui possono disporsi, cinque a cinque, in 30.240 modi (= 10 . 9 . 8 . 7 . 6); sei a sei, in 151.200 modi (= 10 . 9 . 8 . 7 . 6 . 5).

§ **3.** Si dicono *permutazioni* i gruppi in cui entrano ad un tempo tutti gli elementi della serie, disposti in ogni possibile maniera.

Così, se invece di disporre 7 colori in gruppi di due o di tre o di quattro, ecc., li dispongo in gruppi, in cui entrino *tutti sette ad un tempo*, queste disposizioni, che comprendono l'intera serie, assumono il nome speciale di *permutazioni*.

Le permutazioni di 7 colori saranno dunque:

$$P_7 = 7 \,.\, 6 \,.\, 5 \,.\, 4 \,.\, 3 \,.\, 2 \,.\, 1 = 5040.$$

E in generale, le permutazioni di n elementi:

$$P_n = n(n-1)(n-2)\ldots 3 \,.\, 2 \,.\, 1,$$

formola che può scriversi in senso inverso, così:

$$P_n = 1 \,.\, 2 \,.\, 3 \ldots n.$$

L'espressione $1 \,.\, 2 \,.\, 3 \ldots n$, designata spesso più semplicemente col simbolo $n!$ dicesi il *fattoriale* di n.

§ **4.** Si chiamano da ultimo *combinazioni* di m elementi presi due a due, tre a tre ed in generale n ad n, i gruppi che si possono formare riunendo questi elementi due a due, tre a tre ed in generale n ad n, ***in modo però che ciascun gruppo contenga almeno un elemento diverso da quelli di qualsivoglia altro gruppo.***

Il numero delle combinazioni è uguale al numero delle disposizioni diviso per quello delle permutazioni, che possono farsi cogli n elementi di un gruppo. E cioè:

$$^nC_m = \frac{^nD_m}{P_n}.$$

---

(1) Esempi: Le 24 lettere dell'alfabeto, prese tre a tre, dànno luogo a 12.144 (= 24 . 23 . 22) disposizioni. Coi 90 numeri del lotto sono possibili 90 . 89 . 88 = 704 . 880 terni, sempre però considerando come diversi l'uno dall'altro anche i terni formati degli stessi numeri, purchè diversamente disposti.

Infatti nelle disposizioni, dove si considerano come diversi l'un dall'altro anche i gruppi che contengono gli stessi elementi, purchè diversamente disposti, questi gruppi di egual contenuto si ripetono tante volte quante sono le permutazioni possibili dei loro elementi. Così nelle disposizioni di m elementi presi due a due, ogni gruppo binario, per esempio, quello composto di a e di b, si ripete due volte, come ab e come ba, perchè due sono le permutazioni possibili fra i due componenti il gruppo; nelle disposizioni tre a tre, ogni gruppo ternario, per es. quello composto di a, di b e di c, si ripete sei volte (abc, acb, bac, bca, cab, cba), perchè sei sono le permutazioni possibili fra i tre elementi del gruppo. E così via. Dividendo dunque il numero delle disposizioni n ad n per quello delle permutazioni di n elementi, i gruppi simili saranno rappresentati una volta sola e la serie non si comporrà più che di gruppi contenenti almeno un elemento diverso l'uno dall'altro.

Sostituendo a ${}^nD_m$ e a P_n i loro valori, avremo:

$$ {}^nC_m = \frac{{}^nD_m}{P_n} = \frac{m\,(m-1)\,(m-2)\,...\,(m-n+1)}{1\,.\,2\,.\,3\,...\,n}. $$

Esempio: Le quaterne, diverse una dall'altra, che possono formarsi coi 90 numeri della ruota, sono:

$$ \frac{90\,.\,89\,.\,88\,.\,87}{1\,.\,2\,.\,3\,.\,4} = 2.555,190 \text{ (1)}. $$

(1) Si potrebbe facilmente dimostrare che

$$ {}^nC_m = {}^{m-n}C_m. $$

Ma limitiamoci ad un esempio: 10 elementi presi 7 a 7 dànno egual numero di combinazioni, che se fossero presi 3 a 3 (essendo 3 la differenza fra 10 e 7). Infatti:

$$ \frac{10.9.8.7.6.5.4}{1.2.3.4.5.6.7} = \frac{10.9.8}{1.2.3}. $$

Come casi particolari notiamo:

$$ {}^{zero}C_m = {}^mC_m = 1 $$

$$ {}^1C_m = {}^{m-1}C_m = m. $$

La prima formola dice che il numero delle combinazioni di m elementi, presi *zero* a *zero*, è lo stesso di quello che si ottiene prendendo tutti gli m elementi in una volta, ossia si riduce all'unità. Ad esempio:

$$ {}^{zero}C_5 = {}^5C_5 = \frac{5.4.3.2.1}{1.2.3.4.5} = 1. $$

La seconda significa che m elementi, presi ad 1 ad 1, dànno lo stesso numero di combinazioni che se fossero presi a gruppi di $m-1$; quel numero è m.

Esempio:

$$ {}^1C_5 = {}^4C_5 = \frac{5.4.3.2}{1.2.3.4} = 5. $$

§ 5. Fin qui abbiamo considerato elementi di una sola e medesima serie, che in varii modi si associano fra loro. Dobbiamo ora considerare elementi di una serie, che si combinano, non fra loro medesimi, ma con elementi di altra o di altre serie somiglianti.

Siano le due serie:

$$a, b, c, d \ldots$$
$$\alpha, \beta, \gamma, \delta \ldots$$

la prima composta di m elementi, la seconda di n.

È ovvio che uno qualsiasi dei termini della prima serie, per es. a, può combinarsi con ognuno degli n termini della seconda, dando origine ad n gruppi binarii. Ma ciò che si è fatto con a può ripetersi per tutti i termini della prima serie ed essendo questi in numero di m, daranno luogo ad $m \,.\, n$ gruppi binarii.

Se $m = n$, cioè se le serie sono egualmente numerose, i gruppi binarii saranno $m \,.\, m = m^2$.

Se le serie fossero tre, composta la prima di m elementi, la seconda di n, la terza di p, è chiaro che ciascuno degli $m.n$ gruppi binarii, formati con elementi delle prime due serie, potrebbe, in unione coi singoli termini della terza, dare origine a p gruppi ternarii; tutti insieme dunque daranno origine a:

$$m \,.\, n \,.\, p \text{ gruppi ternarii.}$$

Nel caso di serie egualmente numerose, cioè se $m = n = p$, i gruppi ternarii sarebbero:

$$m \,.\, m \,.\, m = m^3.$$

Generalizzando, diremo: *supposte diverse serie, il numero dei gruppi che si possono formare prendendo un elemento da ciascuna, si ottiene moltiplicando il numero dei termini di una serie per quello di ciascuna altra serie.*

Esempio: se io getto un dado esaedrico e uno ottaedrico, segnati rispettivamente coi numeri dall'1 al 6 e dall'1 all'8, le combinazioni possibili di punti saranno $6 \,.\, 8 = 48$.

Gettando tre volte di seguito uno stesso dado esaedrico, le combinazioni ternarie possibili sono $6 \,.\, 6 \,.\, 6 = 216$.

Fra 10 uomini e 8 donne possono trattarsi 80 differenti combinazioni matrimoniali.

TITOLO II.

ELEMENTI DI CALCOLO DELLE PROBABILITÀ.

A) *Teoremi fondamentali.*

Sommario: § 1. Concetto di probabilità. — § 2. Teorema della *probabilità composta.* — § 3. Combinazioni di avvenimenti contrarii in ripetute prove. — § 4. Applicazioni alla statistica. — § 5. Teorema di Bernouilli. — § 6. Concetto dello scarto probabile. — § 7. Probabilità di sorpassare o non sorpassare certi limiti in *m* prove.

§ **1.** Nelle nozioni introduttive fu già accennato alla distinzione del concetto comune o soggettivo di « probabilità » dal concetto oggettivo o matematico. *Probabilità*, nel senso matematico, *è il rapporto tra il numero dei casi favorevoli all'aspettazione di un avvenimento e il numero dei casi possibili.* Condizione sottintesa è che tutti i casi, favorevoli o contrarii, siano o si possano ragionevolmente supporre *possibili in egual grado.*

Senza questa restrizione, nota J. Bertrand, la definizione non avrebbe significato; per qualsiasi evento futuro essendo due le « possibilità », che esso si verifichi o che esso non si verifichi, la « probabilità » si potrebbe credere espressa sempre dalla frazione $^1/_2$. Gli eventi aspettati apparirebbero così tutti egualmente probabili, il che è assurdo. Ciò che fa essere la probabilità ora superiore, ora inferiore a $^1/_2$, è appunto il valore generalmente diverso delle due possibilità, le quali entrano in calcolo con un coefficiente d'importanza proporzionato al numero dei casi, da cui sono rappresentate.

Se un dado è ben costrutto e vien gettato senza artificio che ne faciliti la caduta piuttosto su una che su altra delle sei faccie, è ragionevole ritenere che i numeri scritti su queste abbiano *a priori* eguale possibilità di sortita. Quindi il giuocatore che punta, mettiamo, sul numero 2, ha una probabilità di vincita misurata da $^1/_6$, perchè sei sono le faccie del dado con egual possibilità di sortita, e una sola è la faccia che porta segnato il 2; cioè uno solo, su sei, è il caso favorevole all'aspettazione del giuocatore. Se il dado fosse sospetto di cattiva costruzione, la probabilità in questione sarebbe indeterminata. Per determinarla in maniera almeno approssimativa bisognerebbe gettare migliaia di volte il dado e stabilire la graduatoria di frequenza delle sortite dei diversi numeri. Se su 6000 colpi il punto 1 comparisse 600 volte, il 2 appena 500 volte, il 3 ben 1500 volte, ecc., le probabilità rispettive ad un nuovo colpo potrebbero ritenersi di $^1/_{10}$, di $^1/_{12}$, di $^1/_4$, ecc.

L'importante dunque è di ben enumerare i casi favorevoli e i contrarii all'aspettazione, l'insieme dei quali costituisce i casi possibili.

Getto due dadi in una volta; quale probabilità ho di fare la somma 8? Le combinazioni possibili di punti sono 36, perchè ognuna delle 6 faccie del primo dado può trovarsi dalla sorte associata con una qualsiasi delle 6 del secondo; ma le combinazioni favorevoli alla mia aspettazione sono 5 soltanto, perchè in 5 modi e non più può ottenersi la somma 8; ossia col 6 del primo dado e il 2 del secondo; ovvero col 2 del primo e il 6 del secondo; e ancora col 5 e col 3, ovvero col 3 e col 5; infine col 4 da entrambi i dadi. La probabilità cercata è di $^5/_{36}$. L'apprezzamento di questo risultato non può ragionevolmente variare da individuo ad individuo.

In generale, chiamando a il numero dei casi favorevoli, b quello dei contrarii, si avrà $a + b = \mathrm{N}$, numero dei casi possibili. La probabilità dell'avvenimento sarà espressa da:

$$\frac{a}{a+b} = \frac{a}{\mathrm{N}}.$$

Se $b = o$, cioè se non ci sono casi contrarii, sarà: $\frac{a}{a} = 1$, cioè l'avvenimento è *certo;* se $a = o$, cioè se non ci sono casi favorevoli, si avrà: $\frac{o}{b} = 0$, ossia l'avvenimento è ***impossibile.***

La probabilità contraria all'avvenimento è:

$$\frac{b}{a+b} = \frac{b}{\mathrm{N}};$$

e la somma delle probabilità favorevoli e delle contrarie sarà:

$$\frac{a}{a+b} + \frac{b}{a+b} = 1$$

ossia darà la *certezza;* infatti è certo che l'avvenimento aspettato si verificherà o non si verificherà.

Nell'ordinamento del nostro Lotto, 90 sono i numeri della ruota e 5 i numeri che vengono estratti a sorte.

Chi giuoca un numero come ***estratto semplice*** ha evidentemente 5 casi in suo favore (perchè egli può sperare su ciascuno dei 5 estraendi) su 90 possibili. La probabilità di vincere è per lui $^5/_{90} = {}^1/_{18}$.

Chi giuoca invece un numero, come ***estratto determinato***, ha un solo caso in suo favore su 90 possibili. Probabilità $= {}^1/_{90}$.

Per l'*ambo* i casi favorevoli sono le combinazioni binarie, cui possono dar luogo i 5 estraendi; mentre i casi possibili son tutte le combinazioni binarie formabili coi 90 numeri della ruota. Quindi la probabilità dell'ambo, per chi giuoca due numeri, è:

$$\frac{{}^2C_5}{{}^2C_{90}} = \frac{\frac{5 \,.\, 4}{1 \,.\, 2}}{\frac{90 \,.\, 89}{1 \,.\, 2}} = \frac{5 \,.\, 4}{90 \,.\, 89} = \frac{1}{400{,}5}.$$

Per il *terno* i casi favorevoli sono le combinazioni ternarie, che si possono formare coi 5 estraendi; i casi possibili, tutte le combinazioni ternarie dei 90 numeri della ruota. Ossia la probabilità è:

$$\frac{{}^3C_5}{{}^3C_{90}} = \frac{\frac{5.4.3}{1.2.3}}{\frac{90.89.88}{1.2.3}} = \frac{5.4.3}{90.89.88} = \frac{1}{11.748}.$$

Similmente si troverebbe la probabilità del *quaterno*:

$$\frac{{}^4C_5}{{}^4C_{90}} = \frac{1}{511.038}$$

probabilità assai piccola, pari a quella che si avrebbe di estrarre di primo acchito pallina nera da un'urna che contenesse 511.037 palline bianche e una sola nera (1).

(1) È facile calcolare l'utile lordo (teorico) che lo Stato ritrae dal Lotto nelle diverse sorti, quando si sappia che, in caso di vincita, il premio è di 10 $^1/_2$ la posta per l'estratto semplice, di 52 $^1/_2$ volte la posta per l'estratto determinato, di 250 volte per l'ambo, di 4250 per il terno, di 60.000 volte la posta pel quaterno. Nell'ipotesi di un giuoco perfettamente equo e astrazion fatta dalle spese di amministrazione, lo Stato a chi punta una lira sull'estratto semplice dovrebbe promettere, in caso di vincita, 18 lire di premio, perchè la probabilità di vincere, da parte del giuocatore, è solo di $^1/_{18}$; invece lo Stato dà 10 lire e mezza; la differenza tra 10 $^1/_2$ e 18, cioè lire 7 $^1/_2$, è il suo utile lordo, pari a 41,67 %. (Infatti 7 $^1/_2$: 18 = 41,67 : 100). Similmente a chi punta una lira su un terno secco (cioè giocando tre numeri) lo Stato promette in caso di vincita 4250 lire, invece di 11.748, quante ne dovrebbe promettere nell'ipotesi di un giuoco perfettamente equo e astraendo dalle spese d'amministrazione, perchè il giuocatore ha appena $^1/_{11748}$ di probabilità di vincere; la differenza fra 4250 e 11.748 = 7498 corrisponde ad un utile lordo di 63,82 %. E così via. Nel seguente prospettino diamo, accanto agli utili lordi teorici delle varie sorti, gli utili effettivi medii verificatisi in Italia negli undici esercizi dal 1892-1893 al 1902-1903.

	Utile lordo teorico	Utile medio effettivo
Estratto semplice	41,67 per cento	52,67 per cento
Estratto determinato	41,67 »	38,81 »
Ambo	37,58 »	40,80 »
Terno	63,82 »	64,16 »
Quaterno	88,16 »	89,72 »

E qui non sarà inutile una considerazione. All'alto utile lordo che lo Stato si assicura colla privativa del Lotto fa riscontro, com'è noto, un alto utile netto, quale può dare appunto un'industria monopolistica. Se il giuoco fosse abbandonato alla privata concorrenza, è certo che per effetto di questa gli esercenti lotterie si contenterebbero di un utile lordo e netto minore, discendendo fino al limite dell'ordinario profitto; il che significa che lascierebbero maggior margine di vantaggio ai giuocatori. La passione del giuoco ne sarebbe così alimentata, mentre oggi trova un freno nelle condizioni in cui si esercita la privativa dello Stato.

§ 2. Fondamentale è il teorema della « probabilità composta » che si enuncia così: *La probabilità di un avvenimento dovuto al concorso di altri avvenimenti indipendenti tra loro è uguale al prodotto delle probabilità semplici di questi.*

Sia a' il numero dei casi favorevoli ed n' il numero dei casi possibili per rispetto all'avvenimento A, isolatamente considerato; sarà $\frac{a'}{n'}$ la probabilità semplice del verificarsi di A. Sia poi a'' il numero dei casi favorevoli ed n'' quello dei possibili per rispetto all'avvenimento B; sarà $\frac{a''}{n''}$ la probabilità semplice di B. Si domanda qual'è la probabilità del concorso dei due avvenimenti. Se questi sono *indipendenti*, è chiaro che ciascuno dei casi a' favorevoli al primo evento potrebbe trovarsi associato dalla sorte coll'uno, l'altro o l'altro degli a'' casi favorevoli al secondo; sicchè il prodotto $a' . a''$ ci darà il numero dei casi favorevoli al concorso dei due eventi. Del pari, ognuno dei casi n' possibili rispetto ad A potrebbe trovarsi associato dalla sorte con qualsiasi degli n'' casi possibili rispetto a B; sicchè il prodotto $n' . n''$ ci darà l'insieme dei casi possibili per rispetto all'avvenimento composto di A e di B. La probabilità di questo sarà dunque espressa da:

$$\frac{a' . a''}{n' . n''}.$$

Esempio: In un'urna trovansi 12 palline, di cui 5 bianche e 7 nere; in un'altra urna se ne hanno 15, di cui 2 bianche e 13 nere. La probabilità semplice di estrarre di primo acchito pallina bianca dalla prima urna è $^5/_{12}$; quella di estrarre pallina bianca dalla seconda, indipendentemente dal risultato che si possa avere dalla prima, è $^2/_{15}$. Invece la probabilità di estrarre ad un tempo pallina bianca tanto dalla prima, quanto dalla seconda urna, è:

$$\frac{5}{12} \times \frac{2}{15} = \frac{10}{180}.$$

Infatti 180 sono i casi possibili: che l'una o l'altra od altra delle 12 palline della prima urna si combini coll'una o l'altra od altra delle 15 della seconda ($12 \times 15 = 180$); e 10 sono i casi favorevoli all'aspettazione, cioè che l'una o l'altra delle due bianche della seconda urna si combini con l'una o l'altra od altra delle cinque bianche della prima ($2 \times 5 = 10$).

Se, continuando lo stesso esempio, si domandasse qual'è la probabilità di avere pallina nera da entrambe le urne, visto che le probabilità semplici sono $\frac{7}{12}$ e $\frac{13}{25}$, si avrebbe con identico ragionamento:

$$\frac{7}{12} \times \frac{13}{15} = \frac{91}{180}.$$

E così la probabilità d'avere pallina bianca dalla prima urna e pallina nera dalla seconda è data da:

$$\frac{5}{12} \times \frac{13}{15} = \frac{65}{180}.$$

E quella d'avere pallina nera dalla prima e bianca dalla seconda:

$$\frac{7}{12} \times \frac{2}{15} = \frac{14}{180}.$$

Sommando le singole probabilità composte trovate:

$$\frac{10}{180} + \frac{91}{180} + \frac{65}{180} + \frac{14}{180} = \frac{180}{180} = 1,$$

si ha l'unità, come simbolo di certezza, perchè è ben certo che uno dei quattro avvenimenti dovrà verificarsi.

§ 3. Un'importante applicazione del teorema della probabilità composta riguarda la probabilità delle combinazioni di avvenimenti contrari in ripetute prove ed osservazioni. Diconsi avvenimenti contrari quelli, di cui l'uno può verificarsi solo ad esclusione dell'altro in una sola prova. Il nascituro, ad esempio, sarà maschio o femmina; domani pioverà o non pioverà, ecc.

Sia A un evento avente la probabilità semplice p di verificarsi in una prova od osservazione; sia B l'evento contrario con una probabilità q, e sia $p + q = 1$, l'uno o l'altro dei due eventi dovendo accadere.

In *due* osservazioni o prove potrà darsi che si verifichi A tutte due le volte; oppure che tutte due le volte si verifichi B; oppure nella prima si presenti A e nella seconda B; o infine nella prima B e nella seconda A.

Gli avvenimenti composti e le probabilità rispettive saranno:

A A	$p \cdot p$	$= p^2$
A B, B A	$p \cdot q, \; q \cdot p$	$= 2\, p \cdot q$
B B	$q \cdot q$	$= q^2$

Cioè nel caso di due prove le probabilità delle varie combinazioni sono espresse dai termini di sviluppo del binomio $(p + q)$ elevato alla seconda potenza. Infatti: $(p + q)^2 = p^2 + 2pq + q^2$.

Suppongansi ora *tre* prove od osservazioni. Potrà darsi che A succeda tutte tre le volte di seguito, o che tutte tre le volte si verifichi B, o ancora che A si verifichi due volte e B una (in tre diversi modi di disposizione) o infine che A compaia una sola volta e B due (pure in tre diversi modi di disposizione). Gli avvenimenti composti e le rispettive probabilità saranno:

A A A	$p \cdot p \cdot p$	$= p^3$
A A B, A B A, B A A	$p \cdot p \cdot q, \; p \cdot q \cdot p, \; q \cdot p \cdot p$	$= 3p^2 \cdot q$
B B A, B A B, A B B	$q \cdot q \cdot p, \; q \cdot p \cdot q, \; p \cdot q \cdot q$	$= 3pq^2$
B B B	$q \cdot q \cdot q$	$= q^3$

Ma p^3, $3\,p^2 q$, $3\,p\,q^2$, q^3 non sono che i termini di sviluppo del binomio $(p + q)$ elevato alla terza potenza; per loro mezzo possiamo dunque calcolare facilmente le probabilità delle diverse combinazioni di due avvenimenti contrari in tre prove od osservazioni.

Proseguendo per parità di ragionamento e generalizzando, avremo che le probabilità delle varie combinazioni degli avvenimenti contrari A e B, in m prove od osservazioni, sono espresse dai singoli termini di sviluppo del binomio $(p + q)$ elevato alla *emmesima* potenza:

$$(p + q)^m = p^m + m \, . \, p^{m-1} \, . \, q + \frac{m\,(m-1)}{1 \, . \, 2}\, p^{m-2} \, . \, q^2 +$$

$$+ \frac{m\,(m-1)\,(m-2)}{1 \, . \, 2 \, . \, 3} \, . \, p^{m-3} \, . \, q^3 + \ldots + q^m.$$

Il termine generale di questo sviluppo è, come già sappiamo:

$${}^{x}C_m \, . \, p^{m-x} \, . \, q^x.$$

Facciamo qualche esempio. In Italia su 100 nascite in generale si contano quattro nati-morti e 96 nati-vivi. Si domanda: su tre nascituri (ossia su *tre* prove) qual'è la probabilità che tutti siano vivi o tutti morti o due vivi e uno nato-morto o due nati-morti e uno vivo? L'applicazione della formola dà questi risultati:

Probabilità di 3 *nati-vivi* $= \left(\frac{96}{100}\right)^3 = \frac{884.736}{1.000.000}$

Probabilità di 2 *nati-vivi* e 1 *nato-morto* $= 3 \, . \left(\frac{96}{100}\right)^2 \left(\frac{4}{100}\right) = \frac{110.592}{1.000.000}$

Probabilità di 2 *nati-morti* e 1 *nato-vivo* $= 3 \, . \left(\frac{96}{100}\right) \left(\frac{4}{100}\right)^2 = \frac{4608}{1.000.000}$

Probabilità di 3 *nati-morti* $= \left(\frac{4}{100}\right)^3 = \frac{64}{1.000.000}$

Sommando le diverse probabilità trovate, si ha: $\frac{1.000.000}{1.000.000} = 1$, cioè la certezza; infatti è certo che una delle quattro combinazioni si verificherà.

Ancora. Un'urna contiene 10 palline, di cui 3 bianche e 7 nere. Probabilità di estrazione di una bianca in una prova $= {}^3/_{10}$; di una nera $= {}^7/_{10}$. Si domanda qual'è la probabilità di avere in 6 successive estrazioni (rimettendo ogni volta la pallina sortita nell'urna) 4 volte pallina bianca e 2 volte pallina nera?

Si applichi la formola ${}^{x}C_m \, . \, p^{m-x} \, . \, q^x$, dove $m = 6$; $x = 2$; $p = {}^3/_{10}$; $q = {}^7/_{10}$. Si avrà:

$${}^{2}C_6 \left(\frac{3}{10}\right)^4 \left(\frac{7}{10}\right)^2 = \frac{59.535}{1.000.000},$$

ossia la probabilità cercata è di circa 6 per cento.

§ **4.** L'applicazione di questi primi elementi di calcolo di probabilità alla statistica, si rivela subito feconda di risultati. Vi sono fenomeni d'ordine collettivo che si comportano nelle loro combinazioni e distribuzioni secondo i principii ora esposti di probabilità; e ve ne sono che si scostano invece dalla norma segnata dal calcolo, però in una misura e in una direzione piuttosto costanti. A questa seconda categoria appartengono soprattutto i fenomeni in cui entra come fattore la volontà umana. In entrambi i casi l'utilità del calcolo delle probabilità non può essere disconosciuta, se si pensa che esso solo fornisce in certo modo la pietra di paragone per giudicare di particolari atteggiamenti dei fatti di rilevazione statistica, per classificarli in maniera razionale e metterne in evidenza le leggi empiriche.

Se si trattasse di valutare la probabilità che due nascituri *da due diverse donne* siano entrambi maschi o entrambi femmine o l'uno maschio e l'altro femmina, il teorema della probabilità composta avrebbe la sua diretta applicazione, perchè qui i casi sono certo *indipendenti*. Ma se dovessimo valutare la probabilità delle diverse combinazioni di sesso per due nascituri da una stessa donna (come nella ipotesi di un parto gemello), i casi non si dovrebbero più ritenere *a priori* indipendenti, essendovi ragioni d'ordine biologico, confermate appunto dall'osservazione statistica, per credere che l'analogia di condizioni, in cui si sviluppano gli embrioni umani gemelli, favorisca piuttosto l'identità che non la diversità del loro sesso. Allora il teorema della probabilità composta ci serve solo indirettamente, per stabilire cioè, al confronto dei dati dell'osservazione, *in che misura gli eventi considerati* NON *sono indipendenti*.

Su 100 nascite in Italia si hanno 51,6 maschi e 48,4 femmine. Nei parti gemelli, se gli avvenimenti fossero indipendenti, si dovrebbero avere, in via di probabilità, su 100 casi:

Gemelli		
Entrambi *maschi*	$\left(\frac{51,6}{100}\right)^2$	= 26,63 per cento
Entrambi *femmine*	$\left(\frac{48,4}{100}\right)^2$	= 23,42 »
1 *maschio* ed 1 *femmina*, in due diversi modi di successione	$2 . \left(\frac{51,6}{100}\right)\left(\frac{48,4}{100}\right)$	= 49,95 »
	Totale	100,00 per cento.

Queste le probabilità matematiche. L'osservazione statistica invece ci informa che, su 100 parti gemelli, si hanno:

33,37 gemelli maschi, ossia 6,74 *in più* del calcolo, di cui sopra;
30,16 » femmine, ossia 6,74 *in più* del calcolo;
36,47 » di sesso diverso, ossia 13,48 *in meno* del calcolo.

Il confronto ci fornisce dunque la misura e la direzione in cui spiega la sua efficacia quel complesso di condizioni comuni, in cui si sviluppano gli embrioni gemelli (1).

Ma veniamo a qualcuno dei fatti, in cui la non-indipendenza dei casi è regola generale, in cui cioè interviene come motore la volontà umana.

Veggansi le combinazioni matrimoniali tra persone che sanno o non sanno apporre la propria firma all'atto nuziale.

In Italia, nel 1896, gli sposi e le spose capaci di scrivere il proprio nome furono rispettivamente 63,04 su 100 e 47,44 su 100. Gli analfabeti furono dunque nella proporzione di 36,96 % tra i primi e di 52,56 % tra le seconde. Ora si domanda: qual'è la probabilità che due persone, le quali si sposano, siano entrambe analfabete o lo sia l'una e non l'altra o entrambe sian capaci di apporre la firma all'atto nuziale? Se la coltura primaria e le altre circostanze, che di solito l'accompagnano, fossero elementi indifferenti per la scelta matrimoniale, se insomma gli avvenimenti potessero considerarsi come indipendenti, noi non avremmo che da trattare i rapporti numerici su indicati come espressioni di probabilità semplici e applicare il teorema della probabilità composta, così:

Modi di combinazione	Probabilità
Letterato con letterata	$\frac{63,04}{100} \times \frac{47,44}{100} = 29,91$ per cento
Analfabeta con analfabeta ...	$\frac{36,96}{100} \times \frac{52,56}{100} = 19,42$ »
Letterato con analfabeta	$\frac{63,04}{100} \times \frac{52,56}{100} = 33,13$ »
Analfabeta con letterata	$\frac{36,96}{100} \times \frac{47,44}{100} = 17,54$ »

(1) Questo calcolo è fatto in base ai dati del dodicennio 1884-1895 per l'Italia. In questo stesso periodo si ebbero 1932 parti tripli, di cui 454 di 3 maschi, 460 di 3 femmine, 528 di 2 maschi e 1 femmina, 490 di 2 femmine e 1 maschio.

Calcolando le probabilità matematiche sulla base del rapporto 51,6 % pei maschi e 48,4 % per le femmine, accertato per la natalità in generale, e contrapponendo le percentuali risultanti dall'osservazione statistica, abbiamo:

Parti tripli composti di:	Probabilità matematica	Realtà osservata
3 maschi	13,74 per cento	23,50 per cento
3 femmine	11,34 »	23,81 »
2 maschi e 1 femmina ...	38,66 »	27,33 »
2 femmine e 1 maschio ...	36,26 »	25,36 »

Ma ecco che l'osservazione statistica si scosta non poco dai risultati del calcolo:

COPPIE		Probabilità matematica	Osservazione statistica	Differenze
Gruppi simili..	L^s L^e	29,91 per cento	42,44 per cento	+ 12,53 per cento
	A^s A^e	19,42 »	31,96 »	+ 12,54 »
Gruppi dissimili	L_s L^e	33,13 »	20,60 »	— 12,53 »
	A^s A_e	17,54 »	5,00 »	— 12,54 »

Non vi ha dunque indipendenza o indifferenza di casi; vi ha al contrario una spiccata attrazione o simpatia fra gli individui di gruppi simili, e una repulsione tra quelli di gruppi dissimili. Tra i primi le combinazioni matrimoniali risultano più frequenti che non dovrebbero aspettarsi a calcolo di probabilità; tra i secondi risultano al contrario meno frequenti. In Demografia si dimostra appunto essere questa una legge empirica, che impera nella scelta matrimoniale, per tutti i caratteri di somiglianza o dissomiglianza (età, stato civile, professione, nazionalità, religione, ecc.), intorno a cui si posseggono notizie statistiche.

§ 5. Teniamo ancora presente lo sviluppo del binomio $(p + q)^m$. Il primo termine, p^m, esprime la probabilità che tutte le m volte si verifichi l'avvenimento A e mai l'avvenimento contrario B; il secondo termine, $mp^{m-1}.q$, esprime la probabilità che si verifichi tutte le volte, meno una, A, e una sola volta B; il terzo termine, la probabilità del verificarsi di A tutte le volte, meno due, e del verificarsi di B due volte; e così di seguito, sino all'ultimo termine, q_m, il quale misura la probabilità che si verifichi tutte le m volte B e mai A.

Matematicamente si dimostra (1) che *la combinazione più probabile d'ogni altra è quella nella quale gli avvenimenti si presentano un numero di volte proporzionale alle proprie probabilità semplici;* ossia quella in cui l'evento A si presenta mp volte, e l'evento contrario, B, mq volte (Teorema di J. Bernoulli).

Così nel getto di una moneta, al giuoco detto di « testa e croce », la probabilità semplice della sortita di una delle due faccie, per esempio, *testa*, è $^1/_2$; e la combinazione più probabile su 10 prove è che sorta 5 volte *testa* e 5 *croce;* su 100 prove, che esca 50 volte l'una e 50 l'altra. Nel getto del dado, la probabilità semplice di fare il punto 4 è $^1/_6$; quella di fare un punto diverso dal 4 è $^5/_6$; su 6000 colpi la combinazione più probabile è che il 4 si presenti 1000 volte e gli altri punti, presi insieme, 5000, cioè un numero di volte proporzionale alle rispettive probabilità semplici.

---

(1) Veggasi, per la dimostrazione, J. BERTRAND, op. cit., pag. 71.

Tutte le altre combinazioni possibili hanno una probabilità tanto minore, quanto più grande è il divario tra la combinazione aspettata e quella che si dovrebbe avere in ragione delle probabilità semplici degli avvenimenti.

Così al giuoco sopradetto di « testa e croce » la combinazione più probabile o tipica, su 100 colpi, è che sorta 50 volte *testa* e 50 *croce*; invece la sortita di 51 volte *testa* e 49 *croce* (o viceversa) ha una probabilità alquanto minore; quella di 52 e 48, una probabilità ancor più piccola, e via dicendo; finchè la sortita di 100 volte *testa* e zero volte *croce* (o viceversa) ha una probabilità così infinitesimale, da doversi ritenere l'evento praticamente impossibile, se la moneta è ben costrutta e il modo col quale la si getta non è artificioso.

La misura di queste diverse probabilità si può avere *direttamente* e *rigorosamente* dai termini di sviluppo del binomio; senonchè, in caso di prove molto numerose, i calcoli riuscirebbero praticamente impossibili. Il termine di sviluppo del binomio, che dà la probabilità della combinazione tipica, è:

$$^{mq}C_m \, . \, p^{mp} \, . \, q^{mq}.$$

Se m è molto grande e se si fa $(p + q) = 1$, questo termine è espresso per approssimazione dalla formola:

$$\frac{1}{\sqrt{2 \pi m p q}}$$

dove $\pi = 3{,}14...$; m è il numero delle prove; p e q le probabilità semplici degli avvenimenti aspettati (1).

Così al getto del dado la comparsa di un punto determinato, ad esempio, il 3, per 1000 volte su 6000 tratti — combinazione tipica —, avrebbe la probabilità:

$$\frac{1}{\sqrt{2 \, . \, 3{,}14 \, ... \, 6000 \, . \, \frac{1}{6} \, . \, \frac{5}{6}}} = \frac{1}{72{,}36}.$$

Similmente la comparsa di 50 volte *testa* e 50 *croce* al getto della moneta — combinazione tipica su 100 prove — avrebbe la probabilità di $\frac{1}{12{,}53}$.

Ogni altra combinazione, che si scosti di una quantità x dalla combinazione tipica, ha una probabilità rigorosamente espressa da

$$^{mq+x}C_m \, . \, p^{mp-x} \, . \, q^{mq+x}$$

(1) Questa formola d'approssimazione riposa alla sua volta sulla celebre formola di Stirling:

$$1 \, . \, 2 \, . \, 3 \, ... \, n = n^n \, . \, e^{-n} \sqrt{2 \pi n}$$

dove e (base dei logaritmi naturali) $= 2{,}718...$

e approssimativamente dalla formola:

$$\frac{1}{\sqrt{2\pi m p q}} \cdot e^{-\frac{x^2}{2mpq}}$$

(dove $e = 2{,}718 \ldots$).

La qual formola vale per termini non lontani dal ***maximum***, quando m è assai grande e p e q non sono molto diseguali (1).

Su 100 colpi la probabilità d'avere 58 volte *testa* e 42 *croce*, ossia uno scarto $x = 8$, sarebbe dunque:

$$\frac{1}{\sqrt{2 \cdot 3{,}14 \cdot 100 \cdot \frac{1}{2} \cdot \frac{1}{2}}} \cdot 2{,}718 \ldots^{-\frac{64}{2 \cdot 100 \cdot \frac{1}{2} \cdot \frac{1}{2}}} = \frac{1}{45{,}08}.$$

§ 6. Vi ha uno scarto o scostamento della combinazione tipica, che per la sua particolare importanza merita di essere rilevato. È lo scarto detto ***probabile*** per antonomasia, ossia quella deviazione che presenta eguale probabilità di essere o non essere sorpassata in un dato numero di prove. La sua espressione matematica è:

$$0{,}4769 \ldots \sqrt{2 m p q}.$$

Su 6000 tratti di dado, la combinazione tipica, dicemmo, è che un punto determinato, per esempio, il 3, compaia 1000 volte e gli altri punti, presi insieme, 5000 volte.

Lo scarto o la deviazione probabile è:

$$0{,}4769 \sqrt{2 \cdot 6000 \cdot \frac{1}{6} \cdot \frac{5}{6}} = \pm 19{,}5$$

la qual cosa significa che colui il quale scommettesse che su 6000 colpi il punto 3 sortirà un numero di volte compreso fra 981 e 1020 (o tra 980 e 1019) avrebbe egual probabilità di vincere o di perdere, ossia avrebbe tante probabilità di vincere quante l'avversario suo, il quale scommettesse invece che il punto 3 sortirà un numero di volte compreso fra zero e 980 o fra 1019 e 6000. In altre parole, le 39 combinazioni aspettate dal primo giuocatore possono realizzarsi in tanti e diversi modi, quanti son quelli in cui possono realizzarsi tutte le altre combinazioni immaginabili.

---

(1) Questa condizione a torto è di solito sottaciuta, mentre è chiaro che se ad esempio, p fosse piccolissimo in confronto di q, l'asimmetria della seriazione si farebbe presto sensibile, mentre la formola indicata dà in ogni caso una seriazione simmetrica. — Veggasi a proposito la tabella di valori del Bortkewitsch per l'applicazione della formola di probabilità nel caso di p piccolissimo.

Ciò significa ancora che, se ripetendo molte serie di 6000 tratti di dado ciascuna, risultasse costantemente uno scarto superiore a 20 dalla combinazione tipica, sarebbe giustificato il sospetto di una causa perturbatrice, alterante la probabilità semplice del punto aspettato.

§ 7. La probabilità che su m prove l'avvenimento avente la probabilità semplice p compaia un numero di volte compreso fra

$$mp \pm t\sqrt{2mpq}.$$

è una funzione di t, indipendente da m, da p e da q.

Per i valori di questa funzione si trova un prontuario utilissimo in appendice alle opere già citate del Bertrand e del Ferrero. Noi lo riportiamo qui, ma, per dir così, a sezione molto ridotta.

Scegliamo alcuni valori per meglio far intendere l'applicazione della formola.

In una serie qualunque di prove, si ha la probabilità di $\frac{1}{2}$, ossia si può scommettere 1 contro 1, che l'avvenimento avente la probabilità semplice p comparirà un numero di volte compreso nei limiti:

$$mp \pm 0{,}4769\sqrt{2mpq}.$$

Ci sono invece $^{73}/_{100}$ di probabilità (più precisamente 7.300.104 diecimilionesimi), che esso abbia a presentarsi un numero di volte compreso fra

$$mp \pm 0{,}78\sqrt{2mpq}.$$

Similmente ci sono $^{9}/_{10}$ di probabilità, ossia si può scommettere 9 contro 1, che esso non uscirà dai limiti

$$mp \pm 1{,}163\sqrt{2mpq}.$$

Orbene, per venire alle applicazioni ai fatti collettivi, dato un rapporto statistico qualsivoglia, se le variazioni sue oscillano intorno ai limiti matematici dello scarto probabile, noi potremo affermare con relativa sicurezza che il fenomeno in esame ha una *probabilità costante* e che le deviazioni dalla media, cioè le piccole irregolarità della sua manifestazione quantitativa, dipendono da minute e molteplici influenze accidentali o varianti senza regola assegnabile e non suscettive quindi di analisi statistica, del genere di quelle, per le quali un punto del dado invece di sortire giusto 1000 volte su 6000, sorte 1015 o 1003 o 980, ecc. volte. Se invece, in una serie abbastanza lunga di osservazioni, il rapporto statistico in questione presenta variazioni che sorpassano sempre e notevolmente i limiti dello scarto probabile, allora si ha ragion di credere che tali variazioni dipendano in parte da qualche speciale causa perturbatrice preponderante su tutto l'insieme delle altre cause d'ordine secondario e accidentale.

Uno dei rapporti demografici più costanti è quello, ad esempio, dei sessi nelle nascite. Nel periodo 1883-1902 si ebbero in Italia, in media annuale, 566.965 nati-vivi maschi e 535.941 femmine, ossia $51,_{406}$ maschi e $48,_{594}$ femmine per 100 nati-vivi in generale. Se questo rapporto si assume come espressione di una *probabilità costante* (con che il fatto della produzione dei sessi verrebbe assimilato alla estrazione di palline di due diversi colori da un'urna), vuol dire che si potrà scommettere uno contro uno che in un anno di media frequenza di nascite il numero dei maschi sarà compreso fra i limiti

$$m\,p \pm 0{,}4769 \ldots \sqrt{2\,m\,p\,q},$$

cioè. essendo m (media delle nascite) $=$ 1.102.906; $p = 51,_{406}$ %, $q = 48,_{594}$ %, il numero dei maschi sarà compreso fra 566.965 + 354 e 566.965 — 354, ossia fra 567.319 e 566.611. Ai quali limiti corrispondono i rapporti $51,_{439}$ % e $51,_{374}$ %. Orbene, nel ventennio considerato, il rapporto effettivo dei maschi al totale dei nati-vivi si mantenne per 11 volte nei limiti indicati e per 9 volte li varcò; sicchè, *a priori*, non sarebbe stata cattiva scommessa quella supposta. Ecco i risultati del *Movimento dello stato civile* in Italia:

Anni	Maschi per 100 nati-vivi	Anni	Maschi per 100 nati-vivi
1883	51,463	1893	**51,414**
1884	**51,419**	1894	51,355
1885	51,518	1895	**51,413**
1886	51,468	1896	**51,382**
1887	**51,430**	1897	**51,428**
1888	51,366	1898	**51,392**
1889	51,469	1899	51,300
1890	51,369	1900	51,335
1891	**51,391**	1901	**51,430**
1892	**51,396**	1902	**51,384**

N.B. — Sono indicati in grassetto i rapporti, che si mantennero nei limiti dello scarto probabile.

Noi potremmo ora, col sussidio delle formole accennate e della tavola di valori di t, risolvere altri quesiti. Per rimanere nell'esempio ora proposto, domandiamoci qual era lo scarto, che aveva egual probabilità di essere o non essere sorpassato una volta su venti. Ciò equivale a calcolare i limiti, per cui potevasi scommettere 19 contro uno, che non sarebbero stati oltrepassati in una sola prova annuale. Detti limiti sono espressi dalla formola già nota, in cui il coefficiente t prende il valore di 1,386:

$$m\,p \pm 1{,}386 \sqrt{2\,m\,p\,q} = 566.965 \pm 1029.$$

O in altri termini si avevano $^{19}/_{20} = 95$ % di probabilità che il numero dei maschi su 1.102.906 nascite non avrebbe ecceduto 567.994 e non

sarebbe risultato inferiore a 565.936. Ai quali numeri assoluti fanno riscontro i rapporti $51,_{500}$ % e $51,_{313}$ %. Orbene, secondo il nostro prospetto, solo nel 1885 il rapporto salì a $51,_{518}$ e solo nel 1899 discese a $51,_{300}$, sconfinando cioè due volte dai limiti preveduti, invece di una, ma rimanendo entro tali limiti per 18 volte su 20, se non proprio per 19 volte, come avremmo dovuto avere a calcolo matematico.

La fissità o stabilità del rapporto dei sessi nelle nascite è, come risulta da molt'altre indagini congeneri, veramente meravigliosa; le piccole variazioni si verificano sempre in limiti compatibili coll'ipotesi di una probabilità costante. Ciò significa che la causa o il complesso di cause costanti, da cui dipende la formazione del sesso, è così preponderante, che ogni altra influenza ha carattere di mera accidentalità e neutralizzandosi con altre dello stesso ordine tende a scomparire nella gran massa di casi.

Di molti rapporti statistici, considerati in serie abbastanza lunghe, converrebbe mettere in evidenza il ***grado di stabilità*** o ***instabilità***, confrontando le loro variazioni effettive con quelle assegnabili a calcolo di probabilità. Beninteso che se, insieme alle variazioni di carattere accidentale, ne abbiamo di regolari, periodiche o non periodiche, bisognerà liberare la serie da queste ultime e indi procedere al calcolo. In altre parole, si dovrà prima tentare l'interpolazione e poi, ai diversi stadii di questa, vedere se i nuovi rapporti liberati dalle variazioni presunte regolari oscillano entro limiti compatibili coll'ipotesi di un giuoco di sorte. A rigore la stessa serie, da noi presa in esame, dei rapporti dei sessi nelle nascite per il ventennio 1883-1902, avrebbe dovuto mediante l'interpolazione essere assoggettata ad una depurazione di questo genere, perocchè nel detto periodo, anzi per tutto il periodo che va dal 1872 al 1902, la proporzione dei maschi tra i nati-vivi in Italia ha dimostrato una tendenza piuttosto costante ad attenuarsi, sebbene in piccolissima misura (1).

La tabella a pagina seguente (2) serve, nei casi ordinari, per determinare la probabilità che il fenomeno sia contenuto nei limiti:

$$m p \pm t \sqrt{2 m p q},$$

Gli esempi riportati in questo stesso paragrafo ne chiariscono bene l'uso.

---

(1) Infatti si ebbero in media 106,48 maschi per 100 femmine nel 1872-1875; 106,34 nel 1876-1880; 106,06 nel 1881-1885; 105,85 nel 1886-1890; 105,73 nel 1891-1895 e 105,67 nel 1896-1902.

(2) Essa non è altro che la tavola modificata e semplificata dei valori dell'integrale:

$$\Theta(t) = \frac{2}{\sqrt{\pi}} \int_0^t e^{-t^2} dt,$$

che si trova in tutti i trattati di calcolo della probabilità.

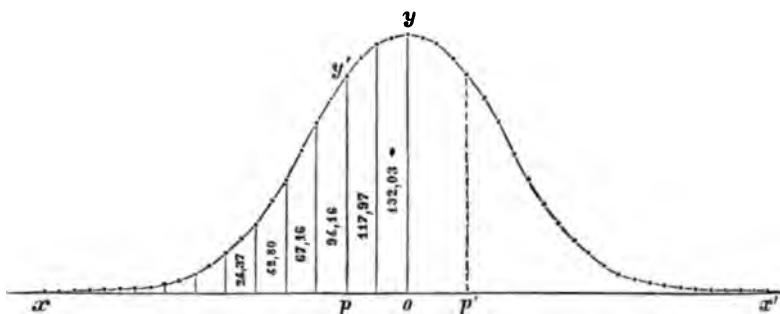
Θ (t)	t	Θ (t)	t	Θ (t)	t	Θ (t)	t
1 %	0,0089	26 %	0,2346	51 %	0,4881	76 %	0,8308
2 »	0,0177	27 »	0,2440	52 »	0,4994	77 »	0,8488
3 »	0,0266	28 »	0,2534	53 »	0,5109	78 »	0,8673
4 »	0,0355	29 »	0,2629	54 »	0,5224	79 »	0,8864
5 »	0,0443	30 »	0,2725	55 »	0,5341	80 »	0,9062
6 »	0,0532	31 »	0,2820	56 »	0,5460	81 »	0,9267
7 »	0,0621	32 »	0,2917	57 »	0,5580	82 »	0,9481
8 »	0,0710	33 »	0,3013	58 »	0,5702	83 »	0,9703
9 »	0,0799	34 »	0,3111	59 »	0,5826	84 »	0,9935
10 »	0,0889	35 »	0,3209	60 »	0,5951	85 »	1,0179
11 »	0,0978	36 »	0,3307	61 »	0,6078	86 »	1,0436
12 »	0,1068	37 »	0,3406	62 »	0,6208	87 »	1,0707
13 »	0,1157	38 »	0,3506	63 »	0,6339	88 »	1,0994
14 »	0,1247	39 »	0,3607	64 »	0,6473	89 »	1,1301
15 »	0,1337	40 »	0,3708	65 »	0,6609	90 »	1,1631
16 »	0,1428	41 »	0,3810	66 »	0,6747	91 »	1,1988
17 »	0,1518	42 »	0,3913	67 »	0,6888	92 »	1,2379
18 »	0,1609	43 »	0,4017	68 »	0,7032	93 »	1,2812
19 »	0,1700	44 »	0,4121	69 »	0,7179	94 »	1,3299
20 »	0,1791	45 »	0,4227	70 »	0,7329	95 »	1,3859
21 »	0,1883	46 »	0,4333	71 »	0,7482	96 »	1,4522
22 »	0,1975	47 »	0,4441	72 »	0,7639	97 »	1,5345
23 »	0,2067	48 »	0,4549	73 »	0,7800	98 »	1,6450
24 »	0,2160	49 »	0,4659	74 »	0,7965	99 »	1,8214
25 »	0,2253	**50** »	**0,476936**	75 »	0,8134	—	—

B) *Principio o legge degli errori accidentali.*

Sommario: § 1. Postulati ed ipotesi. — § 2. Curva degli errori e tavola di frequenza.

§ **1.** Il calcolo delle probabilità presta altro notevole servizio alla statistica col ***principio*** (impropriamente: *legge*) ***degli errori accidentali.*** Quando più osservazioni riferibili ad uno stesso oggetto o fenomeno sono discordanti, ma pur degne di eguale fiducia, la media aritmetica dei risultati fornisce il miglior criterio di scelta o anche il valore più probabile della misura cercata. Ciò implica l'ipotesi *che gli errori, se proprio* ***accidentali, abbiano eguale probabilità di verificarsi così in più come in meno***, così nel senso di esagerare il risultato, come nel senso di impicciolirlo. Una seconda ipotesi, confermata del resto assai

spesso dall'esperienza, è che gli ***errori piccoli siano più facili a commettersi dei grandi***, o in altre parole che la frequenza degli errori sia una certa funzione della loro grandezza. Così misurando il tempo impiegato dal suono per giungere al nostro orecchio da una distanza determinata, accadrà più facilmente di sbagliare di qualche decimo di secondo, che di uno, due o più secondi.

Partendo da questi postulati, il Gauss dimostrò che la frequenza degli errori accidentali in ordine di grandezza è tale da potersi rappresentare con una curva simmetrica e illimitata d'ambo le parti, che si può considerare come il limite cui tende il poligono descritto coi termini di sviluppo del binomio $(p + q)^m$, quando m si faccia sempre più grande e sia $p = q$. La simmetria della curva rispetto all'ordinata centrale è appunto una proprietà derivante dall'ipotesi che errori accidentali, positivi e negativi, della stessa grandezza, siano egualmente probabili; e l'essere poi la curva assintotica o illimitata, nel senso che i suoi due rami si avvicinano sempre più all'asse delle ascisse senza mai toccarlo, significa che un errore, anche grandissimo, è teoricamente possibile, sebbene estremamente improbabile (1).

§ 2. Sull'asse xx' s'intende fissata la scala degli errori o scostamenti dalla media aritmetica assunta come valor più probabile; da o verso x gli errori in meno (o negativi), da o verso x' gli errori in più (o positivi). Le aree comprese fra due ordinate qualunque figurano la frequenza teorica degli errori entro i limiti di grandezza indicati sull'asse xx', ai punti su cui si elevano le ordinate in questione. Nel nostro diagramma, a sinistra, è poi segnato l'intervallo op, il quale delimita colle ordinate oy, py' e colla curva yy' una porzione d'area eguale alla metà della figura oyx, ossia eguale a un quarto dell'intera figura. Questo intervallo dicesi *scarto probabile* per antonomasia. Supposte 1000 osservazioni graduate secondo gli scarti loro dalla media aritme-

---

(1) L'equazione di questa curva è:

$$y = A e^{-k^2 x^2}$$

ossia non è che l'equazione semplificata già riferita al § 5 di questo Titolo, p. 223

tica, lo scarto *op* delimiterebbe quelle 250, i cui risultati, sebbene inferiori alla media, le sono tuttavia più prossimi. Lo stesso scarto, preso a destra del punto *o*, delimiterebbe le 250 osservazioni più prossime alla media, e con risultati superiori. Se prendessimo un intervallo eguale alla metà dello scarto probabile, esso delimiterebbe tanto a destra che a sinistra dell'ordinata centrale 132 osservazioni su 1000 o poco più, com'è indicato in cifre nel diagramma. Così, conosciuto il valore numerico di *op*, si può col sussidio di un prontuario, che si trova in ogni trattato delle probabilità, calcolare quanti casi od osservazioni debbono teoricamente trovarsi compresi fra la media aritmetica e una deviazione qualsiasi dalla media, oppure fra due deviazioni scelte ad arbitrio. Nel seguente prospetto intanto, preso come unità lo scarto o errore probabile, è data la frequenza degli errori per intervalli procedenti di 50 in 50 centesimi dello scarto probabile.

Grandezza degli errori (preso come unità l'errore o scarto probabile)	Frequenza relativa degli errori				Totale	
	Errori in più		Errori in meno			
Da 0 a 0,5	132,03	250	132,03	250	264,07	500
» 0,5 a 1	117,97		117,97		235,93	
» 1 a 1,5	94,16		94,16		188,32	
» 1,5 a 2	67,16		67,16		134,33	
» 2 a 2,5	42,80		42,80		85,60	
» 2,5 a 3	24,37		24,37		48,73	
» 3 a 3,5	12,39		12,39		24,78	
» 3,5 a 4	5,63		5,63		11,26	
Oltre 4	3,49		3,49		6,98	
Totali . . .	500,00		500,00		1000,00	

Pongasi, che misurando mille volte l'altezza di una montagna, l'errore probabile sia risultato di 10 metri. In altri termini, 500 osservazioni abbiano dato uno scostamento (dalla media delle mille misure) inferiore a 10 metri e le altre 500 uno scostamento superiore. Allora, stando ai numeri teorici sopra indicati, un errore di non oltre 5 metri (cioè di non oltre la metà dello scarto probabile) dovette verificarsi in 264 osservazioni su 1000, ossia 132 volte in più e 132 in meno; un errore da 3 a 3 $^1/_2$ volte lo scarto probabile, ossia compreso fra i 30 e 35 metri dovette succedere circa 25 volte (24,78) su 1000; e sette volte soltanto un errore di 40 metri o più (più che quadruplo dello scarto probabile), ecc. Teoricamente, diciamo, le cose dovrebbero essere andate così; la meraviglia è che in molte esperienze la realtà concorda colla teoria: il fatto avvenuto, con quello che il calcolo comportava. Il Bessel, classificando in ordine di grandezza gli errori su 470 osservazioni di

TAVOLA DI FREQUENZA DEGLI ERRORI

m	Numero degli errori su 10.000	*m*	Numero degli errori su 10.000	*m*	Numero degli errori su 10.000	*m*	Numero degli errori su 10.000
0	0	0,62	3242	1,34	6339	2,65	9261
0,01	54	0,64	3340	1,36	6410	2,70	9314
0,02	108	0,66	3438	1,38	6480	2,75	9364
0,03	161	0,68	3535	1,40	6550	2,80	9411
0,04	215	0,70	3632	1,42	6618	2,85	9454
0,05	269	0,72	3728	1,44	6686	2,90	9495
0,06	323	0,74	3823	1,46	6753	2,95	9534
0,07	377	0,76	3918	1,48	6818	3,00	9570
0,08	430	0,78	4012	1,50	6883	3,05	9603
0,09	484	0,80	4105	1,52	6947	3,10	9635
0,10	538	0,82	4198	1,54	7011	3,15	9664
0,12	645	0 84	4290	1,56	7073	3,20	9691
0,14	752	0,86	4381	1,58	7134	3,25	9716
0,16	859	0,88	4472	1,60	7195	3,30	9740
0,18	966	0,90	4562	1,64	7313	3,35	9762
0,20	1073	0,92	4651	1,68	7428	3,40	9782
0,22	1180	0,94	4739	1,72	7540	3,45	9800
0,24	1286	0,96	4827	1,76	7648	3,50	9818
0,26	1392	0,98	4914	1,80	7753	3,55	9834
0,28	1498	**1,00**	**5000**	1,84	7854	3,60	9848
0,30	1604	1,02	5085	1,88	7952	3,65	9862
0,32	1709	1,04	5170	1,92	8047	3,70	9870
0,34	1814	1,06	5254	1,96	8138	3,80	9896
0,36	1919	1,08	5337	2,00	8227	3,90	9915
0,38	2023	1,10	5419	2,05	8332	4,00	9930
0,40	2127	1,12	5500	2,10	8433	4,10	9943
0,42	2230	1,14	5581	2,15	8530	4,20	9954
0,44	2334	1,16	5660	2,20	8622	4,30	9963
0,46	2436	1,18	5739	2,25	8709	4,40	9970
0,48	2539	1,20	5817	2,30	8792	4,50	9976
0,50	2641	1,22	5894	2,35	8870	4,60	9981
0,52	2742	1,24	5971	2,40	8945	4,70	9985
0,54	2843	1,26	6046	2,45	9016	4,80	9988
0,56	2944	1,28	6121	2,50	9082	4,90	9991
0,58	3044	1,30	6194	2,55	9146	5,00	9993
0,60	3143	1,32	6267	2,60	9205	...	...

Bradley relative alle coordinate di una medesima stella, ha verificato che essi seguivano davvicino la *legge* (cosidetta) *degli errori accidentali*. Altri esempi recano i fisici. Gli artiglieri affermano che alla stessa norma obbedisce la dispersione dei colpi al bersaglio intorno al centro di mira. Col Quételet il principio ha trovato applicazioni estese in antropometria e dall'antropometria è passato poi nello studio delle variazioni delle piante e degli animali.

Nel prospetto a pagina precedente è dato, per 10.000 osservazioni, il numero teorico degli errori compresi tra zero e un valore qualunque, preso come unità lo scarto o errore probabile. Di fronte ad $m = 0{,}10$ leggesi 538; ciò significa che su 10.000 errori, 538 non superano il decimo dello scarto probabile, e di essi la metà saranno positivi o in eccesso, l'altra metà negativi o in difetto. Similmente di fronte ad $m = 1{,}50$ troviamo 6883; vuol dire che su 10.000 errori, 6883 non superano una volta e mezza lo scarto probabile, ecc. Se si considerano errori compresi fra due limiti qualunque, basterà calcolare per differenze. Così gli errori di grandezza compresa fra 6 decimi e 7 decimi di scarto probabile, cioè fra $m = 0{,}60$ ed $m = 0{,}70$, sono in numero eguale alla differenza fra 3143 e 3632, cioè 489, di cui metà positivi e metà negativi. Beninteso che per valori di m intermedii fra due indicati nella tabella, il numero degli errori si determinerà per interpolazione.

Se si moltiplicano i valori di m per la costante 0,47693, si ottengono i valori di t, già dati nel prospetto a pag. 227.

Prendiamo ad esempio il confronto fra la distribuzione teorica e quella effettiva degli errori in ordine di grandezza su 100 determinazioni dell'ascensione retta della stella polare, fatte all'osservatorio di Königsberg negli anni 1813-1815. Lo scarto probabile risultò di 0'',88. Se si fa questo scarto eguale all'unità, 0'',4 corrisponderanno in proporzione ad un valore $m = 0{,}453$. Per questo valore di m la nostra tavola indicherebbe, fatta l'opportuna interpolazione, 2400 errori in cifra tonda su 10,000, ossia 24 su 100. Per un valore doppio di m, cioè per $m = 0{,}906$, si hanno 45,9 errori su 100, ossia 21,9 in più. Per un valor triplo di m, cioè per $m = 1{,}359$, se ne hanno 64,1 su 100, cioè altri 18,2 in più, come si vede nella terza colonna del prospetto seguente, e così via.

Errori	Numero effettivo degli errori	Numero calcolato
Da 0'',0 a 0'',4	25	24,0
» 0'',4 a 0'',8	22	21,9
» 0'',8 a 1'',2	19	18,2
» 1'',2 a 1'',6	11	13,7
» 1'',6 a 2'',0	9	9,5
» 2'',0 a 2'',4	8	6,0
» 2'',4 a 2'',8	2	3,4
» 2'',8 a 3'',2	3	1,8
» 3'',2 a 3'',6	1	0,9
Oltre 3'',6	0	0,6

La concordanza fra la teoria e la realtà è soddisfacente.

C) *Tipi di curve di variazioni.*

Sommario: § 1. I cinque tipi di curve *semplici* del Pearson. — § 2. Le curve *composte*.

§ 1. La cosidetta legge degli errori accidentali ha trovato, già dicemmo, per merito del Quételet e di altri (Galton, Weldon, De Vries, Pearson, Duncker) estese applicazioni in antropometria e nello studio delle variazioni delle piante e degli animali. La ripartizione di un gruppo di individui per gradazioni di statura, di peso, di forza misurata al dinamometro, ecc., dà luogo a serie numeriche somiglianti a quella teorica degli errori di osservazione (1). Sicchè, conosciuta la statura media o il peso medio, ecc., dei componenti il gruppo e conosciuto lo scarto probabile, si può determinare *a priori*, con fiducia di allontanarci poco dal vero, il numero di individui compresi fra due limiti, quali si vogliano, di statura o di peso o d'altro. Qui si tratta naturalmente di seriazioni simmetriche.

Il significato di questa legge è che due gruppi di ignote, ma in complesso equivalenti cause elementari partecipano alla formazione delle deviazioni positive e negative del carattere considerato dal suo valor medio. Ogni singola deviazione corrisponde ad una determinata combinazione di cause dell'uno e dell'altro gruppo, e la sua frequenza alla probabilità del concorso di tali cause nel modo supposto fra tutti i modi immaginabili di combinazioni.

Senonchè indagini più larghe, specialmente di statistica zoologica e botanica, hanno provato l'esistenza di *seriazioni asimmetriche*, non assoggettabili alla regola, così semplice e comoda, degli errori accidentali. L'asimmetria talora dipende dalla circostanza che il materiale statistico osservato è eterogeneo e risulta dal miscuglio di due o più materiali omogenei (es. una popolazione mescolata di due razze); tal'altra dal fatto che la tendenza alla deviazione in un dato senso dalla media è più forte della tendenza alla deviazione nel senso opposto. Nel primo caso si tratta di scindere la curva originale nelle sue componenti, problema analiticamente arduo anche nell'ipotesi più semplice della scomposizione in due curve normali, cioè in due seriazioni simmetriche procedenti colla norma degli errori accidentali, ma aventi ciascuna una propria media e un proprio scarto probabile. Il secondo caso richiede che si determini il tipo generale delle curve asimmetriche di frequenza. Già vedemmo che i termini di sviluppo del binomio $(p + q)^m$, graficamente descritti, dànno luogo a curve (limiti di poligoni) tanto più asimmetriche quanto maggiore è la differenza fra p e q.

---

(1) Veggasi un esempio nei nostri *Principii di Demografia*, pag. 93-94, riguardante la ripartizione dei coscritti italiani per statura.

La teoria generale, che le riguarda, fu svolta dal Pearson, ma non può rientrare nei confini del nostro studio (1). Dobbiamo limitarci ad alcuni cenni.

Le curve a una sola *norma* (*unimodal curves*, come le chiamano gli Inglesi) possono essere semplici o composte. Le *semplici* si distinguono nei cinque tipi seguenti:

a) *Limitate da ambo le parti*, cioè tali che i loro rami toccano l'asse delle ascisse a distanze finite dall'ordinata massima. Esse sono *asimmetriche* (Tipo I), ovvero *simmetriche* (Tipo II);

b) *Limitate da una sola parte*, cioè tali che uno dei rami tocca l'asse delle ascisse a distanza finita dall'ordinata massima, mentre l'altro si avvicina indefinitamente all'asse, senza mai toccarlo. Queste curve sono necessariamente *asimmetriche* (Tipo III);

c) *Illimitate da ambo le parti*, cioè tali che i due rami si avvicinano sempre più all'asse delle ascisse senza mai toccarlo. Si distinguono in *asimmetriche* (Tipo IV) e *simmetriche* (Tipo V) (2).

Le equazioni, quali son date dal Pearson, di alcune di queste curve, contenendo parecchie costanti vincolate alla variabile, non rendono chiari, quanto si può desiderare, i confronti di seriazioni diverse. Noi riteniamo che il metodo indicato a pag. 194-195, anche portato ad un ulteriore stadio d'interpolazione, risponda meglio alle ordinarie esigenze della Statistica.

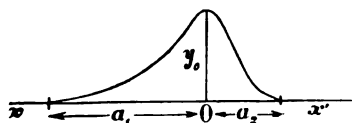
§ 2. Le curve a una sola norma, ma *composte*, possono risultare da due curve semplici con norme differenti, ma così vicine tra loro, da confondersi ancora in una, nella somma che si fa dei termini delle due seriazioni; ovvero dalla somma di due curve aventi la stessa norma, ma un diverso scarto probabile. Il Pearson ha indicato il modo per decomporre una curva composta, avente in apparenza una sola norma,

---

(1) Veggasi Karl Pearson, *Contributions to the mathematical Theory of Evolution*, nelle *Philosophical Transactions of the Royal Society of London*, 1894, volume CLXXXV *A*; 1895, vol. CLXXXVI *A*; 1896, vol. CLXXXVII *A*; 1901, vol. CXCVII *A*.

Si possono consultare con vantaggio pure i seguenti scritti, in cui sono volgarizzate le idee del Pearson: Georg Duncker, *Die Methode der Variationsstatistik*, Leipzig, Engelmann, 1899, e C. B. Davenport, *Statistical Methods with special reference to Biological Variation*, New York, J. Wiley, 1899.

(2) Diamo qui la figura e l'equazione dei diversi tipi di curve:

Tipo I.

$$y = y_0 \left(1 + \frac{x}{a_1}\right)^{\nu a_1} \left(1 - \frac{x}{a_2}\right)^{\nu a_2}.$$

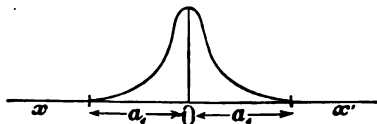
L'origine delle x è al punto o dove si eleva l'ordinata massima y_0. I tratti

in due con due distinte norme; ma il procedimento è assai laborioso e complicato.

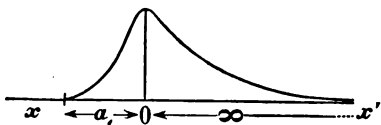
Il Livi trattò il problema in modo empirico, sufficiente però a darne un'idea alle persone non famigliari colle matematiche, nelle sue ricerche sui miscugli etnici (1). Egli ha supposto una mescolanza, in parti eguali, di due popolazioni della statura media di m. 1,63 la prima, di m. 1,69 la seconda, ed aventi lo stesso scarto probabile. In tali condizioni la curva del miscuglio differisce dalle due primitive per la cima più pianeggiante e la base più larga. Questi caratteri della curva composta si accentuerebbero, se si supponesse il miscuglio di due popolazioni, l'una della statura di m. 1,60, l'altra di m. 1,72. Ma affinchè la curva

---

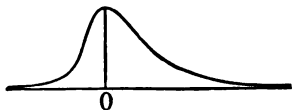
limitati sull'asse delle x, ai quali si estendono i due rami della curva, sono a_1 e a_2. La curva diventa immaginaria quando $x = -a_1$, $x = a_2$.

Tipo II.

$$y = y_0 \left(1 - \frac{x^2}{a_1^2}\right)^{\nu a_1}.$$

L'equazione di questa curva, simmetrica e limitata d'ambo le parti, deriva da quella del tipo I, essendo $a_2 = a_1$.

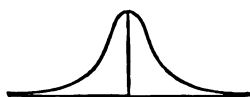
Tipo III.

$$y = y_0 \left(1 + \frac{x}{a_1}\right)^{\nu a_1} . e^{-\nu x}.$$

L'equazione di questa curva, illimitata nella direzione di x^1, deriva da quella del tipo I, quando si faccia $a_2 = \infty$. Il simbolo e (base dei logaritmi naturali) ha il valore di 2,718...

Tipo IV.

$$y = y_0 \cos \theta^{2m} . e^{-\nu\theta},$$

dove $\operatorname{tang} \theta = \frac{x}{a}$.

È il tipo più comune di curve oblique biologiche; l'equazione però contiene elementi trigonometrici che la rendono poco chiara nell'uso ordinario.

Tipo V.

$$y = y_0 e^{-hx^2}.$$

L'equazione di questa curva, nota sotto il nome di *curva normale* o *degli errori accidentali*, deriva da quella del tipo II, quando si faccia a_1 grandissimo.

(1) *Sulla interpretazione delle curve seriali in Antropometria* (*Atti della Società romana di Antropologia*, Roma 1895).

del miscuglio possa presentare una traccia dei vertici delle due curve componenti, occorre una differenza ancor maggiore nelle medie, ad es., di 15 o 16 cm., che nessuna popolazione europea presenta nei grossi gruppi etnici, di cui si compone, almeno per il carattere « statura ». Per l'indice cefalico ciò potrebbe darsi; se mettessimo insieme 10,000 piemontesi o romagnoli, che sono i più brachicefali tra gli italiani e 10,000 sardi, che sono i più dolicocefali, la curva del miscuglio presenterebbe traccia dei due vertici delle componenti. Essa dovrebbe classificarsi adunque tra le *curve plurinormali* (*multimodal curves*, degli autori inglesi).

A parte il caso di una pluralità di norme apparenti (che possono risultare da scarsezza di osservazioni o da una suddivisione in classi troppo minuta per rispetto al numero delle osservazioni), le curve

plurinormali possono avere le loro norme separate da seni più o meno profondi. Il rapporto tra la depressione esistente fra le due norme e l'altezza della norma minore si chiama *indice di isolamento*. E si chiama *indice di divergenza* la distanza tra le norme espressa in termini dello scostamento medio della più variabile fra le curve componenti. In biologia queste curve indicano bene spesso la condizione polimorfica di una specie; tale il caso del numero dei fiori a raggio della margheritina bianca, che ha norme ai valori 8, 13, 21, 34, ecc. presi sull'asse delle ascisse (Davenport). In generale il significato delle curve composte è che non più due soli gruppi (eguali o ineguali) di cause ignote ed operanti in senso positivo l'uno e negativo l'altro, producono le variazioni del carattere intorno ad uno stesso valor normale; ma che esistono più coppie di cotali gruppi di cause, ciascuna coppia svolgendo la propria azione in un determinato circolo di individui. Conseguenza di ciò è il differenziarsi di una forma già unitaria in più forme o tipi distinti (Duncker).

Dal poco che abbiam detto si comprenderà l'importanza del metodo seriale:

1° nel determinare il *tipo* in un complesso di varietà individuali;

2° nel precisare in che misura la tendenza alla deviazione in un certo senso dalla media supera la tendenza alla deviazione nel senso opposto;

3° nel rendere facilmente confrontabili le seriazioni di uno stesso fenomeno collettivo osservato in diversi tempi o luoghi, o di più fenomeni affini, bastando all'uopo paragonare fra loro le costanti caratteristiche delle equazioni delle curve.

Quanto al primo punto, è chiaro che la media aritmetica, se può mettere in risalto l'elemento tipico, quando la seriazione è simmetrica,

non raggiunge lo scopo allorchè la seriazione è asimmetrica. Il tipo è indicato piuttosto da quel valore, cui corrisponde il massimo addensamento dei soggetti osservati. Il secondo punto è d'interesse speciale nelle questioni di biologia. L'asimmetria delle curve di variazioni degli organi delle piante e degli animali può significare che le condizioni di esistenza e di riproduzione per gli individui variati in una certa misura e direzione dalla norma, sono più facili (o più difficili) di quelle cui sono soggetti gli individui variati in senso opposto; sicchè il tipo stesso, per lenta selezione, non sarebbe fisso, ma tenderebbe a spostarsi in una direzione determinata.

D) *La « legge dei piccoli numeri »*.

Sommario: § 1. Formola di probabilità per fenomeni rari e tavola del Bortkewitsch. — § 2. Applicazioni statistiche.

§ **1.** Il primo tentativo di trattare le serie di piccoli numeri, cioè i fenomeni rari, dal punto di vista del calcolo delle probabilità, è dovuto a L. von Bortkewitsch (1). Comunemente si crede che i piccoli numeri in Statistica sfuggano ad ogni disciplina, la compensazione delle cause accidentali aspettandosi solo nelle grandi masse. Però se pensiamo che la « rarità » di un avvenimento implica per ciò stesso la « frequenza » dell'avvenimento contrario e che l'una e l'altra sono teoricamente vincolate nella stessa formola e praticamente dipendono da un dato concerto di condizioni oggettive, esteriori, non ci deve far meraviglia il trovare che, come il fenomeno più comune segue davvicino la norma segnata dal calcolo, così la segua il fenomeno più raro e per dir così « complementare ». Se il numero dei casi, in cui si presenta il primo, concorda abbastanza con quello assegnabile a calcolo, anche il numero dei casi, che rimangono per esaurire la serie delle possibilità, e nei quali comparirà il secondo fenomeno, non potrà non concordare col rispettivo numero teorico.

Sia p la probabilità che si verifichi l'evento A, e $q = 1 - p$ quella che non si verifichi, in una prova. La probabilità che A compaia x volte su m prove od osservazioni è espressa da:

$$\frac{m\,(m-1)\ldots(m-x+1)}{1\,.\,2\,.\,3\ldots x}\,p^x\,.\,q^{m-x}.$$

Se m è grandissimo e p piccolissimo, per modo che $mp = k$ sia costante, l'espressione precedente ha per limite:

$$\frac{k^x\,.\,e^{-k}}{1\,.\,2\,.\,3\ldots x}.$$

(1) Cit. a pag. 22.

La formola suppone k piccolo in confronto di m (1). Col sussidio di essa il Bortkewitsch preparò una tabella, in cui per determinati valori di k sono indicate le probabilità che l'avvenimento non compaia affatto o compaia 1 volta, 2 volte, x volte in una unica serie di m osservazioni o prove. Diamo qui solo uno *specimen* di questo utile prontuario:

VALORI DI $\dfrac{k^x \, . \, e^{-k}}{1 \, . \, 2 \, . \, 3 \ldots x}$ ESPRESSI IN DECIMILLESIMI.

x	$k = 0,1$	$k = 0,2$	$k = 0,3$	$k = 0,4$	$k = 0,5$	$k = 0,6$	$k = 0,7$	$k = 0,8$	$k = 0,9$
0	9048	8187	7408	6703	6065	5488	4966	4493	4066
1	905	1637	2223	2681	3033	3293	3476	3595	3659
2	45	164	333	536	758	988	1217	1438	1647
3	2	11	33	72	126	198	284	383	494
4	—	1	3	7	16	30	50	77	111
5	—	—	—	1	2	4	7	12	20
6	—	—	—	—	—	—	1	2	3

(1) Infatti, essendo

$$p = \frac{k}{m}, \text{ e } q = 1 - p = 1 - \frac{k}{m},$$

il termine generale di sviluppo del binomio diventerà:

$$\frac{m(m-1)\ldots(m-x+1)}{1 \, . \, 2 \, . \, 3 \ldots x} \left(\frac{k}{m}\right)^x \left(1 - \frac{k}{m}\right)^{m-x}$$

e tenendo presente che, per m grandissimo,

$$\left(1 - \frac{k}{m}\right)^m = e^{-k},$$

si ha con facili trasformazioni:

$$\frac{m(m-1)\ldots(m-x+1)}{1 \, . \, 2 \, . \, 3 \ldots x} \cdot \frac{k^x}{(m-k)^x} \, . \, e^{-k}$$

che può anche scriversi:

$$\frac{m}{m-k} \cdot \frac{m-1}{m-k} \ldots \frac{m-x+1}{m-k} \cdot \frac{k^x \, . \, e^{-k}}{1 \, . \, 2 \, . \, 3 \ldots x}.$$

Ma per m assai grande e k piccolo, ognuno dei fattori $\frac{m}{m-k}$, $\frac{m-1}{m-k}$, ecc., è approssimativamente eguale all'unità; quindi anche il loro prodotto sarà all'incirca eguale a 1. La formola si riduce quindi a:

$$\frac{k^x \, . \, e^{-k}}{1 \, . \, 2 \, . \, 3 \ldots x}.$$

(*Segue*) VALORI DI $\frac{k^x \cdot e^{-k}}{1 \cdot 2 \cdot 3 \ldots x}$ ESPRESSI IN DECIMILLESIMI.

x	$k=1,0$	$k=1,2$	$k=1,4$	$k=1,6$	$k=1,8$	$k=2,0$	$k=2,2$	$k=2,4$	$k=2,6$
0	3679	3012	2466	2019	1653	1353	1108	907	743
1	3679	3614	3452	3230	2975	2707	2438	2177	1931
2	1839	2169	2417	2584	2678	2707	2681	2613	2510
3	613	867	1128	1378	1607	1804	1966	2090	2176
4	153	260	395	551	723	902	1082	1254	1414
5	31	62	111	176	260	361	476	602	735
6	5	12	26	47	78	120	175	241	319
7	1	2	5	11	20	34	55	83	118
8	—	—	1	2	5	9	15	25	38
9	—	—	—	—	1	2	4	7	11
10	—	—	—	—	—	—	1	1	3
11	—	—	—	—	—	—	—	—	1

x	$k=2,8$	$k=3,0$	$k=3,2$	$k=3,4$	$k=3,6$	$k=3,8$	$k=4,0$	$k=4,5$	$k=5,0$
0	608	498	408	334	273	224	183	111	67
1	1703	1494	1304	1135	984	850	733	500	337
2	2384	2240	2087	1929	1771	1615	1445	1125	842
3	2225	2240	2226	2186	2125	2046	1954	1687	1404
4	1557	1680	1781	1858	1912	1944	1954	1898	1755
5	872	1008	1140	1264	1377	1477	1563	1708	1755
6	407	504	608	716	826	936	1042	1281	1462
7	163	216	278	348	425	508	595	824	1044
8	57	81	111	148	191	241	298	463	653
9	18	27	40	56	76	102	132	232	363
10	5	7	13	19	28	39	53	104	181
11	1	1	4	6	9	13	19	43	82
12	—	—	1	2	3	3	6	16	34
13	—	—	—	—	1	1	2	6	13
14	—	—	—	—	—	—	1	2	
15	—	—	—	—	—	—	—	1	2

Per il modo di servirsi di questa tabella valgano brevi chiarimenti.

Supponiamo che un fenomeno d'una certa rarità si presenti in un paese colla media di 2 casi all'anno. Questo numero non dice per se stesso nulla sulla probabilità semplice del fenomeno, la quale potrà

essere dell'1 per 1000 se l'ambiente, in cui esso si verifica, ammette 2000 casi possibili:

$$\left(2 = 2000 \times \frac{1}{1000}\right),$$

ovvero di 5 centomillesimi, se l'ambiente comporta 40.000 possibilità:

$$\left(2 = 40.000 \times \frac{5}{100.000}\right), \text{ ecc.}$$

Ma a noi basta assumere la media di 2 casi verificati, come valore approssimato di k, ossia come il prodotto approssimativamente noto di due fattori ignoti, quali sono m e p. Orbene la tabella del Bortkewitsch alla colonna $k = 2,0$, indica che il non verificarsi nemmeno di un caso dell' avvenimento in questione in un dato anno ha la probabilità di 0,1353; il verificarsi di un caso solo ha la probabilità 0,2707; il verificarsi di 5 casi, la probabilità 0,0361, ecc. Beninteso che se gli anni considerati fossero 10, 20 o 30 invece di un solo, bisognerebbe moltiplicare per 10, per 20 o per 30 le probabilità anzidette; i prodotti esprimeranno il numero probabile di anni vuoti di casi o segnalati da un caso, due casi, ecc., del fenomeno raro.

§ 2. In Prussia nel venticinquennio 1869-1893 si ebbero 60 *suicidii di fanciulli* men che decenni d'ambo i sessi, ossia una media di 2,4 casi all'anno. Dei 25 anni in questione, 3 non ebbero a registrare alcuno di questi suicidii eccezionali; altri 3 ne registrarono 1 per ciascuno; altri 9 ne registrarono 2 e così via, come leggesi qui sotto. Ora, moltiplicando per 25 i valori indicati nella tabella, alla colonna $k = 2,4$ si ottiene il numero teorico d'anni in cui si sarebbero dovuti aspettare 0, 1, 2, ecc., casi di suicidio di fanciulli.

Numero dei suicidii di fanciulli	N° degli anni in cui fu osservata la cifra controindicata di suicidii	N° degli anni in cui dovevasi aspettare la cifra controindicata di suicidii
0	3	2,3
1	3	5,4
2	9	6,5
3	4	5,2
4	4	3,1
5	1	1,5
6	1	0,6
7 e più	—	0,4
	25	25,0

La concordanza della realtà col calcolo è appena sufficiente, causa la brevità del periodo considerato; ma vien meglio in evidenza, se si raggruppano i dati due a due:

Numero dei suicidii di fanciulli	Anni di osservazione	Anni a calcolo
0-1	6	7,7
2-3	13	11,7
4-5	5	4,6
6 e più	1	1,0
	25	25,0

Un esempio ora di casa nostra: i *parti tripli* nell'Umbria nei 35 anni del periodo 1868-1902. Essi furono in complesso 90; media annua 2,57.

Le probabilità corrispondenti al valore $k = 2,57$ non sono specificate nella tabella nostra; ma essendo questo valore di k compreso fra 2,4 e 2,6, sarà facile ottenerle per interpolazione lineare. Moltiplicando per 35 le probabilità così calcolate, avremo il numero teorico d'anni, in cui erano da aspettarsi 0 casi, 1 caso, 2 casi, ecc., di parti tripli nel compartimento in questione.

Numero dei parti tripli	N° degli anni in cui si osservò la cifra controindicata	N° degli anni in cui era da aspettarsi la cifra controindicata
0	4	2,68
1	6	6,88
2	7	8,84
3	10	7,58
4	5	4,87
5	1	2,50
6	—	1,07
7	1	0,39
8	1	0,13
9	—	0,04
10 e più	—	0,02
	35	35,00

Raggruppando due a due:

x	Anni osservati	Anni calcolati
0-1	10	9,56
2-3	17	16,42
4-5	6	7,37
6-7	1	1,46
8 e più	1	0,19
	35	35,00

Nel raggruppamento dei dati due a due l'accordo della osservazione colla teoria diviene soddisfacente. Ma senza entrare in più minuti particolari, basti dire che con una opportuna combinazione di serie di piccoli numeri, concernenti uno stesso fenomeno raro, osservato, per esempio, in diversi luoghi o presso diverse classi di persone, il Bortkewitsch ha ottenuto concordanze singolari. Anche nei piccoli numeri adunque si apre allo statistico un campo di interessanti esplorazioni.

CAPO QUINTO

Tecnica e logica dei numeri-indici.

TITOLO I.

NUMERI-INDICI SEMPLICI.

Sommario: § 1. Le espressioni sintetiche di fatti complessi. — § 2. Indici di *variabilità* di un gruppo. — § 3. Indici di *preferenza* o *attrazione* nella scelta matrimoniale. — § 4. Indici di *regressione*. — § 5. Indici di *ripartizione dei redditi*, ecc. — § 6. Indici delle *variazioni dei prezzi*.

§ 1. I *numeri-indici* rispondono al bisogno di avere per certi fatti complessi le espressioni più sintetiche e insieme più comode per confrontare i risultati di osservazioni fatte in tempi e luoghi diversi o in categorie di fenomeni affini. L'ideale dello statistico, come già dicemmo, dev'essere di ridurre un *maximum* di materiale numerico greggio ad un *minimum* di elementi caratteristici e decisivi, su cui fissare le idee.

Quando il fatto o fenomeno, nonostante la sua estrinseca complessità, obbedisce ad una legge semplice, le *costanti* dell'equazione colla quale possiamo rappresentare la serie o seriazione, servono esse medesime da numeri-indici. Così è un fatto complesso la ripartizione dei redditi in un paese ad epoche differenti o per più paesi in uno stesso tempo; è un fatto anche più complesso il combinarsi degli individui nelle nozze secondo la somiglianza o dissomiglianza di età, di professione, di religione, di coltura, ecc.; tuttavia con facili processi arriviamo a districare dal materiale greggio gli elementi necessari per la rappresentazione semplificata del fenomeno e per la immediata intelligenza dei confronti internazionali. Altre volte invece, meglio che un fatto specifico, possiamo avere un insieme di fatti specifici, cui corrisponde una generalizzazione della nostra mente, una sintesi d'impressioni, che pur vorremmo tradurre in indici numerici. Nel linguaggio ordinario corrono spesso frasi di questo genere: « il benessere economico del paese è in aumento » — oppure « il movimento intellettuale sembra arrestarsi » — oppure « la moralità è in decadenza », ecc. Senza dubbio, allorquando ci esprimiamo in tal guisa, noi raccogliamo in una le impressioni ricevute da una serie di fatti distinti, testimonianze in favore o contro un dato asserto, testimonianze che possono anche non essere tutte concordi, nè avere egual grado di attendibilità o egual peso. Per sentenziare, poniamo, intorno alla moralità pubblica, i fatti semplici e distinti, che sono il substrato della nostra generalizzazione, saranno forse: il crescere o

diminuire di certe forme di delinquenza, il crescere o diminuire della natalità illegittima, delle separazioni coniugali o dei divorzi; il crescere o diminuire degli scioglimenti di consigli comunali per malversazioni e disordini amministrativi, delle elezioni politiche contestate per brogli, ecc. Similmente un giudizio sintetico sul movimento intellettuale di un paese non può non tener conto della frequenza degli allievi nei vari gradi di scuole, della frequenza dei lettori nelle biblioteche, delle percentuali degli sposi o dei coscritti analfabeti, del numero delle opere pubblicate, della corrispondenza postale, ecc.

Orbene, anche in questi casi si è tentato dagli statistici di tradurre in numeri-indici e secondo norme razionali e fisse le situazioni complesse, sottraendole in certa guisa all'apprezzamento oscillante, arbitrario degli individui e alla elastica imprecisione del linguaggio comune. Siamo, dico, ancora allo stadio dei tentativi, ma questi non sono indegni di un cenno.

Distinguiamo dunque due specie di numeri-indici: quelli che servono a caratterizzare fatti complessi ma specifici; e quelli che si riferiscono ad una pluralità o combinazione di fatti specifici. I primi li diremo *numeri-indici semplici*, i secondi *numeri-indici composti*.

§ **2.** *Indici di variabilità di un gruppo.* — Quando gli individui componenti un gruppo abbastanza omogeneo si ripartiscono per gradazioni di un carattere determinato secondo la norma degli errori accidentali, riesce facile segnare il campo di variabilità del gruppo, ragguagliando lo *scarto probabile* alla *media* del carattere considerato. Il rapporto, tradotto in percentuale, serve da numero-indice.

In Italia, secondo osservazioni fatte sui coscritti, la statura media dei ventenni è di 1 metro e 631 mm.; lo scarto probabile, di 43 mm.; il che significa (non sarà inutile rammentarlo) che fra 1 metro, 588 mm. (= 1,631 — 43) e 1 m., 674 mm. (= 1,631 + 43) era compresa la metà dei soggetti misurati, l'altra metà ripartendosi in parti eguali fra le stature superiori a 1,674 e quelle inferiori a 1,588. Ora il rapporto fra lo scarto probabile e la media:

$$\frac{43}{1631} = 2{,}64\ \%$$

è precisamente l'*indice di variabilità* del gruppo osservato, per quanto concerne il carattere « statura ».

Questo semplice metodo ci pone in grado di far confronti interessanti, paragonando fra loro seriazioni di individui appartenenti a razze o famiglie etniche diverse.

Evidentemente, se due gruppi d'individui hanno, poniamo, la stessa statura media, ma uno scarto probabile diverso, il gruppo che presenta lo scarto minore ci apparirà più uniforme, più addensato attorno alla propria media in confronto di quello che ha lo scarto maggiore. Se poi

i due gruppi differiscono anche per la media, quello che presenta uno scarto più piccolo in rapporto alla media stessa, si dovrà dire il più uniforme e il meno variabile. L'indice di variabilità può così servire di prezioso criterio per decidere se una popolazione è più o meno pura di un'altra, se è più o meno commista di elementi divergenti dal tipo che predomina.

Gli arruolati italiani, senza distinzione di regioni, hanno un indice cefalico medio di 82,7, con uno scarto probabile di 3,25. Il rapporto

$$\frac{3,25}{82,7} = 3,93\ \%$$

dà la misura della variabilità loro e quindi del loro addensamento intorno alla media. Invece gli arruolati sardi, distintamente considerati, presentano un indice cefalico medio di 77,5, con uno scarto probabile di appena 2,5. La metà di essi si raccoglie dunque intorno alla media fra limiti più ristretti, che non gli italiani in genere. L'indice di variabilità si riduce a

$$\frac{2,5}{77,5} = 3,23\ \%$$

e permette di concludere che la popolazione sarda è probabilmente più tipica, più omogenea della italiana in generale.

Un medesimo gruppo di individui può risultare molto addensato intorno al valor medio per certi caratteri; poco addensato, e cioè avente un campo esteso di variabilità, per certi altri. Anche l'osservazione comune ci mostra che gli uomini sono, ad esempio, assai meno variabili o dissimili l'un dall'altro per statura, che non per peso, per forza o per portata visiva.

INDICI DI VARIABILITÀ.

	Uomini	Donne
Statura	2,53 per cento	2,56 per cento
Peso	7,00 »	9,00 »
Forza	9,12 »	14,42 »
Capacità vitale	11,20 »	13,80 »
Portata visiva	16,00 »	21,75 »

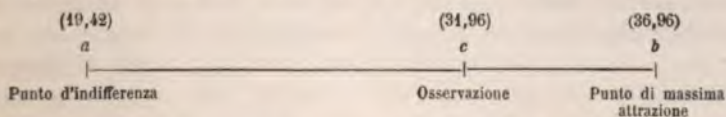
Le ricerche del Galton in argomento, sebbene poco estese, non lasciano dubbi. Gli indici da noi calcolati sui dati dell'autore inglese mostrano pure che la donna è, fisicamente almeno, più variabile, ossia meno tipica dell'uomo: la qual cosa può significare, che i caratteri più divergenti dalla media, esposti nell'uomo a cause particolari di eliminazione, riescono a conservarsi nella donna, protetti da un ambiente di vita più tranquillo (1).

(1) Veggasi *Principii di Demografia*, pag. 101-103 e 119-125.

§ **3.** *Indici di attrazione o preferenza nella scelta matrimoniale.* — Già abbiamo veduto un caso speciale di una regola generale, cioè che gli individui di gruppi simili si preferiscono tra loro nelle combinazioni matrimoniali, mentre quelli di gruppi dissimili si respingono — o, in altri termini, dànno luogo ad un numero di matrimoni inferiore a quel che si potrebbe aspettare a calcolo di probabilità. Ora, vogliamo anche qui esprimere sinteticamente con indici appropriati tali situazioni complesse di cose.

Riprendiamo l'esempio svolto a pag. 220-221. Là abbiamo veduto che nell'ipotesi di combinazioni come in un giuoco di sorte, i matrimoni tra analfabeti e analfabete avrebbero dovuto essere 19,42 su 100; invece la statistica ne registrò 31,96 su 100. Vi fu dunque mutua preferenza o simpatia tra gli individui simili per la mancanza di istruzione primaria. La preferenza deve dirsi tanto maggiore, quanto più la cifra reale eccede quella assegnata dal calcolo. Se l'attrazione fosse stata la massima possibile, se cioè ogni uomo analfabeta non avesse voluto impalmare altra donna che non fosse essa medesima analfabeta, le coppie formate sarebbero state 36,96 su 100. Infatti, secondo l'esempio citato essendo 36,96 su 100 gli sposi analfabeti e 52,56 su 100 le spose pure analfabete, il massimo numero di coppie di analfabeti con analfabete risultava determinato dall'elemento disponibile in quantità minore, cioè dai 36,96 uomini analfabeti.

Ora, col sussidio dei seguenti tre dati: 1° Numero delle coppie che si sarebbero avute a calcolo di probabilità, cioè nell'ipotesi che la somiglianza o la dissomiglianza di caratteri non avesse influito sulla scelta; 2° Numero delle coppie accertate dall'osservazione statistica; 3° Numero delle coppie che si sarebbero avute nell'ipotesi di una attrazione o preferenza massima possibile, si arriva a stabilire un *indice* numerico, che misura il grado di effettiva attrazione.

Rappresentiamo infatti con una retta il tratto di escursione fra il *punto d'indifferenza* (19,42) dato dal calcolo di probabilità, e quello di *maximum* (36,96); il numero reale delle coppie formatesi (31,96) cadrà assai più vicino al punto di *maximum*, da cui dista di sole 5 unità, che non a quello d'*indifferenza*, da cui dista 12,54 unità.

Ogni coppia in più delle 19,42, date dal calcolo, sta a dimostrare che preferenza c'è stata tra gli individui dei due gruppi simili. Allora ragguagliando il tratto d'escursione ac (la cui lunghezza è 31,96 — 19,42 = 12,54 unità) all'intero tratto di escursione ab (la cui lunghezza è 36,96 — 19,42 = 17.54), il rapporto misurerà il grado d'attrazione effettiva, che si è manifestata fra i due gruppi.

L'indice di preferenza o attrazione è questo stesso rapporto ridotto a percentuale:

$$\frac{31,96 - 19,42}{36,96 - 19,42} = \frac{12,54}{17,54} = 71,5\ \%.$$

Con questo metodo abbiamo calcolato nei nostri *Principii di Demografia* gli indici della preferenza che si accordano tra loro nella scelta matrimoniale gli individui di gruppi simili per età, per stato civile, per luogo di nascita, per nazionalità, per professione, per religione. Ne è risultata una graduatoria che indagini più estese e metodi più rigorosi potranno modificare alquanto, ma che non è priva d'interesse. Da essa si apprende che la somiglianza d'età spiega un'influenza più debole nella scelta matrimoniale che non la somiglianza di stato civile; e questa, una più debole della somiglianza di professione; la quale a sua volta è superata dalla somiglianza di coltura, di nazionalità e di religione. Quest'ultima è elemento di tanta importanza nella scelta matrimoniale, che in qualche caso speciale l'indice arriva al 100 %; tale il caso degli israeliti e dei *Confessionlose* a Vienna (che son poi generalmente israeliti), i quali si sposano tutti tra loro.

Caratteri di somiglianza	Indici di preferenza
Età	16,6-19,0 per cento
Luogo di nascita	20,7-31,3 »
Stato civile	31,5-33,1 »
Professione	41,4 »
Istruzione elementare . . .	45,9-71,5 »
Nazionalità	58,0-66,5 »
Religione	21,2-84,9 »

Per la retta interpretazione di questi indici giova tener presente, che i caratteri in esame spiegano un'influenza, che non è tutta loro propria, ma in parte è dovuta ad altri caratteri di somiglianza, celati, per così dire, dietro quelli. Così gli analfabeti scelgono di preferenza la sposa tra le donne analfabete, ma non certo in grazia dell'analfabetismo; bensì piuttosto, perchè dietro questa apparenza o forma di manifestazione esistono ben altri caratteri comuni: somiglianza di condizioni economiche, professionali, relazioni di vicinato, ecc. L'analogia di stato civile implica insieme somiglianza d'età; i vedovi, che passano a seconde nozze, sono generalmente più innanzi negli anni dei celibi e le vedove più delle nubili; e poichè gli individui d'età matura si preferiscono tra loro — o almeno sono ridotti a doversi preferire tra loro — così l'attrazione che ci appare sotto le forme di somiglianza di stato civile è invece, in parte almeno, effetto di somiglianza d'età. Occorrerebbero statistiche assai specializzate per poter sceverare la parte precisa d'influenza che spetta ai singoli caratteri e alle svariate combinazioni loro.

§ 4. *Indici di regressione.* — Trattando delle « correlazioni » ci occorse di accennare alla *legge di regressione* del Galton in materia di eredità dei caratteri. La regressione verso la media è però un fatto generale, che si presenta ogni qualvolta un gruppo scelto di soggetti o di casi può per incrocio o combinazione con elementi di una varietà più lata dare origine a nuovi gruppi o presentarsi sotto nuovi aspetti. La cosa s'intenderà bene per via d'esempi. Formo un gruppo scelto dei piccolissimi comuni nei quali in un dato anno si verificò una forte eccedenza di nascite femminili sulle maschili; indi ricerco, per gli stessi comuni e per anni anteriori o successivi a quello considerato, la proporzione dei sessi nelle nascite. Probabilmente non troverò ripetuta la eccedenza notevole delle femmine se non per qualche comune del gruppo, laddove per altri troverò verificata invece una deficienza più o meno grande oppure un rapporto vicino al normale. Nel complesso si avrà dunque un ritorno, una regressione verso la media. Del pari, se tengo conto separato delle merci d'importazione che da un dato anno al successivo sono rincarite, poco probabile è che al terzo anno della serie le stesse merci risultino tutte rincarite una seconda volta; più probabile invece che alcune di esse siano rimaste stazionarie ed altre rinvilite, sì che nell'insieme si avrà regressione. Formiamo ancora un gruppo dei più giovani tra gli sposi di un dato anno; troveremo che essi hanno scelto spose in media men giovani di loro. Infatti la loro scelta potè spaziare per tutte le classi d'età delle nubili (dai 15 anni in su) e delle vedove; e la preferenza, che molti hanno accordato certo alle donne più giovani, è in parte neutralizzata dalla preferenza che alcuni possono aver avuto per donne di età maggiore della loro. Nell'insieme dunque si deve aver regressione, nel senso che l'età delle spose scelte non è quella che si avrebbe se i più giovani tra gli uomini avessero impalmato unicamente le più giovani tra le donne; e non è nemmen quella che si avrebbe, se i più giovani tra gli uomini avessero preso a caso le loro compagne senza riguardo veruno all'età; ma è intermedia tra queste due.

Valga un ultimo esempio, che illustrerò con una curiosa ricerca. Formando un gruppo scelto dei candidati che ottennero il massimo dei punti in una certa materia d'esame è ovvio che nell'insieme delle altre materie essi avran riportato o riporteranno qualche punto meno del massimo, perocchè accanto alla probabilità di prove *egualmente* (ma non *più*) fortunate, hanno quella di prove *men* fortunate. Per analoga ragione formando un gruppo a parte dei meno quotati in quella materia, facilmente avverrà che essi figurino nel complesso delle altre con qualche cosa più del minimo. Tra questi due gruppi estremi si disporranno gli altri gruppi nell'ordine che loro spetta e con una regressione più o meno accentuata verso la media. Ora il prospetto, che segue, riguardante 178 esaminati nelle 18 materie della Facoltà giuridica di Pavia, mette bene in evidenza cotesta regressione

e graficamente si può rappresentare col metodo tenuto dal Galton a proposito della correlazione fra le stature dei genitori e quelle dei figli loro adulti. La materia d'esame scelta come termine di confronto sono le *Istituzioni di Diritto civile.*

Punti riportati in *Istitux. di Diritto civile*	Media dei punti nelle altre materie	Differenze	Punti riportati in *Istitux. di Diritto civile*	Media dei punti nelle altre materie	Differenze
30	28,10	— 1,90	23	23,86	+ 0,86
29	27,13	— 1,87	22	23,10	+ 1.10
28	27,20	— 0,80	21	22,77	+ 1.77
27	26,13	— 0,87	20	20,87	+ 0.87
26	25,31	— 0,69	19	21,47	+ 2.47
25	24,01	— 0,99	18	21,29	+ 3.29
24	24,09	+ 0,09	...	...	...

Interpolando una retta col metodo dei minimi quadrati e chiamando con x i punti riportati in più o in meno della media 24 nella materia speciale e con y i punti riportati in media nel complesso delle altre materie, si arriva all'equazione:

$$y = 24,_{256} + 0,_{6}\,x.$$

È ovvio che, se il coefficiente di x fosse nullo, la regressione verso la media risulterebbe massima; se quello fosse eguale all'unità, la regressione sarebbe nulla. Nella fattispecie, la regressione effettivamente avvenuta può essere espressa dalla differenza (1 — 0,6) = 0,4.

Quanto più accentuato è il ritorno verso la media, tanto maggiore deve dirsi l'indifferenza colla quale il gruppo scelto subisce le più opposte condizioni dell'ambiente in cui viene a trovarsi. Perciò, anche sotto questo rispetto, l'elaborazione e la comparazione dei dati statistici possono instradarsi per vie promettenti.

§ **5.** *Indici della ripartizione dei redditi, dei patrimoni e simili.* — Di questi indici, che costituiscono un bell'esempio dell'estrema semplificazione, a cui si possono ridurre notevoli masse di dati specialmente nei confronti internazionali, abbiamo detto a proposito dei procedimenti d'interpolazione.

§ **6.** *Indici delle variazioni dei prezzi.* — Da un anno all'altro i prezzi di molte merci possono essere cresciuti, in misure svariatissime; altre merci invece possono essere rinvilite pure in diverse misure, ovvero essere rimaste stazionarie. Posto che si abbiano i dati concreti, si domanda se *in generale* vi fu aumento o diminuzione dei

prezzi. cioè, quale è stata la variazione nel potere d'acquisto della moneta per rispetto alla categoria di beni di cui si tratta.

Il procedimento sarebbe facile se tutti i prodotti considerati avessero, dal punto di vista dell'economia nazionale, la stessa importanza. Basterebbe infatti sommare le loro variazioni percentuali di prezzo, positive o negative o nulle che siano, e dividere il risultato per il numero dei termini dati. Ma se questi non hanno la stessa importanza, il procedimento supposto non regge più. Sarebbe infatti assurdo compensare un rincaro del 10 % nel carbon fossile o nel grano o nel cotone greggio, con un deprezzamento del 10 % negli aghi da cucire, nell'inchiostro o nello zafferano. Bisogna allora per ogni merce istituire il calcolo del valore totale che essa rappresenta *ai prezzi unitari dell'anno che si considera* e del valore totale che essa rappresenterebbe, *se i prezzi unitari*, invece di essere quelli che effettivamente sono, *fossero quelli dell'anno antecedente* o di qualunque altro anno che si assuma come termine di paragone. È implicito dunque che si conosca la quantità fisica dei prodotti, cui si applica il calcolo.

Il metodo torna specialmente acconcio nel caso di merci d'importazione o di esportazione.

Esempio: il paese A ha introdotto dall'estero in un dato anno 1.000.000 di tonnellate di carbon fossile al prezzo unitario di lire 25; 30.000 quintali di canapa a lire 80 per quintale; 4000 capi di bestiame bovino a lire 400 per capo. Nell'anno antecedente i prezzi unitari furono: pel carbone lire 22, per la canapa 85, per il bestiame 400, ossia da un anno all'altro si ebbe per la prima merce un rincaro di 13,6 %; per la seconda un deprezzamento di 5,9 %; per la terza una condizione di cose stazionaria.

Confrontisi ora la spesa effettiva incontrata dal paese nell'acquisto di tali prodotti colla spesa che esso avrebbe incontrata nell'ipotesi che i singoli prezzi fossero rimasti invariati dall'anno antecedente.

	Quantità	Prezzi attuali	Spesa effettiva	Prezzi dell'anno antecedente	Spesa ipotetica secondo i prezzi dell'anno antecedente
		lire	lire	lire	lire
Carbone . . . . *tonn.*	1.000.000	25	25.000.000	22	22.000.000
Canapa . . . . *quint.*	30.000	80	2.400.000	85	2.550.000
Bestiame . . . . *capi*	4.000	400	1.600.000	400	1.600.000
		Totale	29.000.000	*Totale*	26.150.000

Facendo eguale a 100 la seconda somma, la prima diventa eguale a 110,9; il che significa i prezzi essere in media aumentati da un anno all'altro di 10,9 %.

Con siffatto procedimento abbiamo calcolato i numeri-indici all'importazione e all'esportazione per tutti i prodotti indistintamente, che

entrarono nel commercio estero dell'Italia durante il periodo 1882-1902, fatto eguale a 100 lo stato dei prezzi nel 1884 (1).

ANNI	INDICI DEI PREZZI		ANNI	INDICI DEI PREZZI	
	alla importazione	alla esportazione		alla importazione	alla esportazione
1882	112,31	110,30	1893	87,19	86,60
1883	107,32	106,79	1894	81,17	82,45
1884	**100,00**	**100,00**	1895	79,27	83,63
1885	94,12	95,71	1896	79,72	79,09
1886	92,21	95,35	1897	79,07	77,55
1887	90,84	90,87	1898	83,67	79,17
1888	91,25	86,90	1899	87,64	86,45
1889	93,08	89,80	1900	94,90	85,94
1890	93,84	92,94	1901	88,28	83,22
1891	89,71	86,72	1902	84,98	84,26
1892	87,90	86,81	—	—	—

Il metodo tuttavia non è rigoroso, perchè le merci che da un anno all'altro son cresciute di prezzo, dando luogo di solito a minori contrattazioni, entrano in calcolo con un peso o coefficiente d'importanza inferiore a quello che loro altrimenti spetterebbe; laddove il contrario succede per le merci deprezzate. E i due errori non si compensano, ma si sommano. Per evitare tale inconveniente si suggerisce di tener conto per ogni merce non della quantità importata od esportata nell'anno in questione, ma di quella importata od esportata in media annua durante un certo periodo. Ciò peraltro rende laboriosi i calcoli, senza aggiungere molto alla precisione dei risultati.

Il confronto delle due serie di prezzi-indici all'importazione e alla esportazione, permette di stabilire in quali periodi il commercio estero è riuscito più o meno vantaggioso al paese. Allorchè i prezzi delle derrate di esportazione rinviliscono più rapidamente di quelli delle derrate d'importazione — e non vi sia motivo di spiegarne il ribasso coi progressi tecnici delle industrie nazionali, cogli abbondanti raccolti, ecc. — è ovvio che il paese ha comperato caro e venduto a buon mercato o, in altri termini, che per procurarsi la stessa quantità fisica di merci forestiere ha dovuto dare una quantità fisica di merci proprie, maggiore di quella che altrimenti avrebbe ceduta. Il di

(1) Gli indici son calcolati confrontando i valori *provvisori* (cioè dell'anno antecedente) attribuiti alle merci importate od esportate da gennaio a tutto novembre (11 mesi) coi valori *definitivi*, cioè dell'anno di cui propriamente si tratta. Veggasi *Statistica del commercio speciale d'importazione e di esportazione*, pubblicata mensilmente dalla Direzione generale delle Gabelle.

più che ha dato in cambio costituisce una perdita secca. Questo per noi si è verificato nel quinquennio 1888-1892, allorchè tra le difficoltà di trovare nuovi sbocchi a compenso del perduto mercato francese (in seguito alla rottura commerciale colla Francia avvenuta nel 1888), le nostre derrate d'esportazione, che costituiscono il mezzo normale di pagamento dei prodotti e dei capitali importati, si offersero a prezzi ridotti, intanto che i prezzi dei generi d'importazione si mantenevano alti. Aggiungasi che, in quel turno di tempo, grande fu la massa di titoli collocati all'estero per colmare gli sbilanci commerciali e il corso depresso di tali titoli aggravò la perdita secca incontrata dal paese. Un secondo periodo di prezzi più bassi all'esportazione che all'importazione si ebbe tra il 1896 e il 1901, ma certo senza perdita per la nazione, perchè concomitante di innegabili progressi tecnici (riduzioni di costo), di un aumento notevole nella quantità delle merci esportate e fors'anco di un movimento più attivo di forestieri in Italia.

La tecnica dei prezzi-indici lascia ancor molto a desiderare. È discutibile se più convenga tener conto dei prezzi all'ingrosso o di quelli al minuto, dei prezzi delle materie prime e ausiliarie e degli strumenti o di quelli soltanto dei prodotti pronti o quasi pronti al consumo. Il pericolo di duplicazioni indebite nel calcolo difficilmente si può scansare. Quando, per esempio, una manifattura rincarisce, perchè rincarita è la sua materia prima, dovremo introdurre in calcolo entrambi i rialzi di prezzo o uno solo? E se il prezzo all'ingrosso ha sùbito una variazione, mentre i prezzi al minuto sono rimasti invariati, come si formerà l'*index-number?* Ancora: le variazioni di valore di un immobile, che non cambia possessore se non in media ogni 30 anni e più, si sommeranno *sic et simpliciter* con quelle di un titolo, che può cambiar di mano molte volte in un anno? Noi accenniamo solo a tali questioni, non le risolviamo; ma esse lasciano intendere che diverso sarà il procedimento, secondo il punto di vista dal quale si considerano le cose; il punto di vista del consumatore o quello del produttore e del commerciante.

Gli indici dei prezzi, che furono elaborati da molti statistici ed economisti (Sauerbeck, Palgrave, Giffen, Soetbeer, ecc.), non ci possono dare che le variazioni del potere d'acquisto della moneta limitatamente a categorie più o meno ristrette di merci. Le variazioni di valore della moneta, in senso generale, richiederebbero per essere misurate un calcolo esteso a tutto ciò che ha prezzo, a tutto ciò che si scambia in un paese, inclusi quindi gli immobili, i titoli di credito e i servigi personali.

TITOLO II.

NUMERI-INDICI COMPOSTI.

Sommario: § 1. Indici ***unici*** della situazione economica di un paese. — § 2. Condizioni del sistema dell'indice ***plurimo*** o ***composto***. — § 3. Schema rudimentale di indice composto. — § 4. Questioni speciali. — § 5. Metodo del confronto tra le equazioni delle serie. — § 6. Conclusione.

§ **1.** Passiamo ora ad una classe speciale di ***indici***, coi quali si tende a dare un valore numerico alle testimonianze, che ci vengono da diversi fatti e a ricavare da una opportuna elaborazione e composizione di tali testimonianze un giudizio sintetico, un criterio obbiettivo, sottratto all'apprezzamento mobile e vario dei singoli individui.

Sia da esprimere in indici numerici la situazione economica di un paese. Essa si riflette nei modi più svariati nel movimento dei consumi, nella intensità del traffico ferroviario, nell'affluenza maggiore o minore dei depositi presso le Banche, nella proporzione più o meno alta dei protesti rispetto alle cambiali create, nel corso favorevole o contrario dei cambi, ecc.; e persino nella frequenza dei reati contro la proprietà, dei suicidii per dissesti finanziari, dei pegni non riscattati ai Monti di Pietà, dei biglietti giocati al lotto, e così via. Ma la concordanza perfetta delle diverse testimonianze non si ha mai; le deposizioni talvolta sono contradditorie; ovvero concordano nell'indicare la ***direzione*** del movimento economico e discordano nella ***misura:*** in generale non hanno nè egual grado di attendibilità, nè la stessa importanza, nè la stessa chiarezza di significato. Ond'è che la ***semiologia*** (da σημεῖον, segno, sintomo, indizio) va costituendo un capitolo sempre più interessante del metodo statistico.

Alcuni scrittori furono o sono favorevoli al sistema dell'***indice unico*** o ***semplice***, tratto da un fenomeno esteso, ma specifico.

Secondo le idee mercantiliste d'un tempo, l'***abbondanza di moneta metallica*** era misura e segno di benessere, quasi che la ricchezza di un paese dovesse trovarsi in una costante proporzione colla quantità di oro ed argento che serve a farla circolare e distribuire. In contrario si osserva che nei paesi più ricchi e progrediti il sistema degli ***chèques*** e le Stanze di compensazione e l'uso di biglietti fiduciari emessi da grandi Banche tendono a limitare il più possibile gli strumenti metallici della circolazione, che sono più costosi.

Nè meno illusorio è il giudicare della situazione economica di un paese dalla ***popolazione***, col ritenere sintomo favorevole il crescere, sfavorevole il diminuire di essa. Ciò può essere vero in generale, ma può non essere vero in casi particolari. La Francia ha visto triplicarsi la sua ricchezza nel corso del secolo XIX, mentre il numero dei suoi

abitanti crebbe assai più lentamente che in altri paesi; i nostri vicini ritengono anzi necessaria ad assicurare il loro benessere la limitazione della fecondità coniugale.

Il rapporto tra la popolazione improduttiva e il totale della popolazione sarebbe pure, secondo il Ferraris, un buon indice delle condizioni economiche dei paesi. Presso i popoli ricchi il rapporto deve essere alto, perchè minore vi è il bisogno di assoggettare precocemente i fanciulli al lavoro o di ritardare il collocamento a riposo dei vecchi o di sfruttare la donna nel lavoro dei campi e delle officine; il contrario dicasi per le nazioni povere. Ma è molto dubbio se l'interpretazione dei termini « produttivo » e « improduttivo » sia uniforme nei censimenti dei vari paesi, e la rilevazione dei dati egualmente attendibile; d'altronde, siccome i censimenti si ripetono per lo più ad intervalli decennali, non potremmo formare indici delle variazioni dello stato economico, anno per anno, ciò che appunto più interesserebbe.

Un buon indice sarebbe quello del *consumo*, come sosteneva l'Engel; senonchè i rilievi statistici lo lumeggiano ancora scarsamente (1). Conosciamo anno per anno le variazioni nel consumo di alcune derrate, come il sale, il caffè, lo zucchero, il grano, il tabacco, il petrolio, ecc.; ma si ignora quel che è avvenuto per molti altri generi di prima necessità e voluttuari. Qui poi importerebbe aver notizia della parte che prendono ai diversi consumi le diverse classi della popolazione, mentre la raccolta dei bilanci famigliari, massime per le classi lavoratrici, è appena ai suoi inizi e non può colmare le grandi lacune delle nostre osservazioni in argomento.

Critiche di varia specie si possono infine muovere ad altri indici, in quanto si assumano come *unici;* tale quello del *risparmio*, del grado di *agglomerazione delle famiglie nelle abitazioni*, ecc. Il Juglar vede rispecchiate le variazioni dello stato economico di un paese nei *bilanci delle grandi Banche;* ma anche questa è una faccia sola del poliedro (2). Noi siamo quindi condotti a cercare di utilizzare una molteplicità varia di indici, componendola ad unità, per modo che ciascuno funzioni da correttivo di ogni altro o integri il senso proprio di ogni altro.

---

(1) *La Consommation comme mesure du bien-être des individus, des familles et des nations* (*Bulletin de l'Institut international de Statistique*, Tome II, 1re Livraison, 1887).

(2) Clément Juglar, *Des crises commerciales et de leur retour périodique en France, en Angleterre et aux Etats-Unis*, Paris, Guillaumin, 1889. L'autore non nega però valore ad altri indici, ma li considera come sussidiarii di quello desunto dai bilanci delle grandi Banche. Veggasi anche: *Influence des crises commerciales sur l'état économique* (*Journal de la Société de Statistique de Paris*, luglio e settembre 1896), ove il Juglar tien conto speciale degli indici della nuzialità, natalità e mortalità.

§ 2. È merito del Neumann Spallart (1) l'aver tentato di sistemare questo genere di ricerche, distinguendo tre classi di sintomi o indici: *primari* (fenomeni di produzione), *secondari* (scambi, prezzi, salari, ecc.) e *riflessi* (scioperi, emigrazioni, nuzialità...) e riducendoli a serie numeriche comparabili da paese a paese. A dir vero però, la comparabilità delle serie rimaneva un *desideratum*, sia perchè la scelta degli indici proprii di ciascun paese non andava esente d'arbitrio, sia perchè l'elaborazione tecnica e logica delle serie appariva rudimentale. L'idea tuttavia non meritava di cadere in dimenticanza e fu coltivata dal Pantaleoni, dal Coletti e da altri.

Il sistema dell'indice plurimo, se ha da essere qualche cosa più di una semplice curiosità statistica, deve evidentemente soddisfare a due condizioni: 1° avere *una base scientifica*, cioè presentarsi come sviluppo ragionato di certe premesse; 2° avere *possibilità di attuazione*. Indugiamoci alquanto su questi due punti.

Riguardo al primo, poichè ad ogni situazione economica del paese corrisponde un certo grado del suo *benessere*, noi cercheremo le espressioni soggettive di questo e le manifestazioni esteriori od oggettive nei molteplici fatti della vita sociale. Soggettivamente il benessere è sentito come pregio che si attribuisce alla vita, come attaccamento alla patria, come inclinazione ad atti di liberalità, come rispetto alla proprietà altrui, come fiducia e sicurezza dell'avvenire, come somma di soddisfazioni domandabili all'amore, al giuoco, ai consumi di lusso, ecc. Ond'è che il variare del numero dei suicidii, massime per dissesti finanziari, il variare dell'emigrazione, delle donazioni, dei reati contro la proprietà, del risparmio e di altri atti di previdenza, il variare della nuzialità, dei proventi del lotto, del caffè, dello zucchero, del tabacco, ecc., assurgono a manifestazioni quantitative di una serie di stati d'animo presso un popolo. Ogni mutamento economico provoca una modificazione nelle sensazioni di benessere e nella condotta esteriore, diversa nei diversi gruppi, di cui la massa si compone, e diversa secondo il posto che presso ciascun gruppo occupa in un dato istante un dato sentimento nella scala o graduatoria dei sentimenti. Così, nel piccolo àmbito di coloro, pei quali la vita ha già un mediocre valore, l'aggravarsi di una crisi economica moltiplicherà i suicidii — e sarà questa la manifestazione caratteristica per quel gruppo, mentre sarebbe vano ricercarla, ad esempio, nella restrizione dei consumi, nell'astensione dal giuoco, nella rinunzia al matrimonio od altro; — invece nella cerchia di coloro, cui già non ripugnava troppo l'idea di attentare alla altrui proprietà, la supposta crisi moltiplicherà i furti e non già i suicidii; e in quella di coloro che già erano dubbiosi circa la conve-

(1) *Mesure de variations de l'état économique et social des peuples* (*Bulletin de l'Institut international de Statistique*, Tome II, 1re Livraison).

nienza di formarsi una famiglia, moltiplicherà le astensioni dal matrimonio, non già i suicidii, i furti od altro.

Ora è chiaro che se persistessimo a voler ritrovare in un solo indice, per quanto importante, l'espressione e la misura delle condizioni economiche di un intero popolo, faremmo opera necessariamente incompleta e unilaterale; cioè daremmo l'espressione e la misura di stati d'animo particolari per quel solo gruppo, presso il quale un dato sentimento o bisogno occupa il *primo* (o, secondo i casi, l'*ultimo*) posto nella graduatoria. Teoricamente dunque il sistema dell'indice plurimo è logico, anzi è il solo logico.

Resta a vedere se sia attuabile. Noi crediamo che lo sia, pur non nascondendoci le grandi difficoltà dell'impresa.

§ 3. Cominciamo dalla forma più rudimentale, che si potrebbe dare al sistema, salvo a dire poi dei perfezionamenti, ond'è suscettiva.

Suppongasi di non avere, all'infuori dei censimenti, altra rilevazione statistica che quella della *nuzialità*. Niente statistiche economiche o finanziarie, niente statistiche dell'emigrazione, della delinquenza, ecc.

Si siano avuti per un milione di abitanti 8134 matrimoni in un dato anno, 8306 nel successivo, 7738 nel terzo. Ragguagliando a 100 il dato del secondo anno, risulterebbe la serie:

97,9; 100; 93,2.

In mancanza di indici migliori, il movimento economico del paese, il benessere o malessere suo, se così piace dire, si troverà rispecchiato nelle variazioni della nuzialità. Se non si può provare che esse siano imputabili a costumi mutati, allo scoppio di una guerra o ad una deficienza straordinaria di nascite di 25-30 anni addietro, per cui si abbia oggi scarso contingente di giovani nell'età preferita per le nozze, bisognerà ammettere che la nuzialità (di cui non si può negare una certa dipendenza dalle condizioni economiche degli individui) è un indicatore prezioso perchè l'unico, secondo l'ipotesi, statisticamente accertato, del benessere o del malessere nazionale.

Supponiamo adesso di avere a disposizione anche un'altra statistica, quella, ad esempio, del *consumo del caffè*. Da essa risulti che negli anni in questione ogni abitante consumò in media 497, 561, 480 grammi di tale derrata. Ragguagliando a 100 la cifra dell'anno di mezzo, si ha la serie:

88,6; 100; 85,5.

Qualora anche le oscillazioni di detto consumo non si possano spiegare per via di gusti cambiati, di fortunati succedanei, ecc., bisognerà consentire che esse riflettono cangiamenti avvenuti nella ordinaria capacità d'acquisto dei consumatori.

Padroni di due indici, possiamo sommarli anno per anno e prenderne la media aritmetica come *indice composto*. In questa addizione non vi è nulla di illogico. Non è che si addizioni il consumo del caffè colla nuzialità, cose troppo eterogenee, ma gli indici del benessere o malessere, quali li argomentiamo dall'andamento dei matrimoni e dal consumo di quella derrata. Se si bada al primo indice, esso testimonierebbe che quanto a benessere si stava prima al livello di 97,9, rispetto all'anno fisso che segna 100, e che nel terzo anno si è ricaduti a 93,2; se si bada all'altro indice, esso attesta che nel primo e nel terzo anno si stava anche peggio, in confronto dell'anno di mezzo. Diamo *per ora* egual peso alle due testimonianze; la media aritmetica degli indici esprimerà approssimativamente la situazione relativa dei tre anni.

Indici della *nuzialità*	97,9	100	93,2
» del *consumo del caffè* . .	88,6	100	85,5
Media (Indice composto)	**93,25**	**100**	**89,35**

Si conosca adesso anche la *situazione finanziaria dello Stato* e sia l'entrata effettiva rispetto alla spesa effettiva nel rapporto di 97,9 : 100 nel primo anno; di 93,1 : 100 nel successivo; di 81,3 : 100 nell'ultimo. L'ingrossare del disavanzo, segno patologico della situazione, sarebbe espresso da indici decrescenti così:

105,2; 100; 87,3

essendosi per uniformità ragguagliato ancora a 100 il rapporto proprio dell'anno di mezzo. Allora l'indice composto risulterà:

Indici della *nuzialità*	97,9	100	93,2
» del *consumo del caffè* . .	88,6	100	85,5
» della *situazione finanziaria*	105,2	100	87,3
Media (Indice composto)	**97,23**	**100**	**88,67**

Evidentemente l'errore che si commette nel rappresentare la situazione relativa di cose da un anno all'altro, tende a farsi tanto più piccolo, quanto maggiore è il numero e l'accordo delle testimonianze, sempre sottintendendosi che siano soddisfacentemente risolte le questioni che riguardano l'elaborazione tecnica, il significato e il peso o coefficiente d'importanza da assegnarsi ad ogni indice.

Innanzi di trattare tali questioni, importa tener presente che una condizione essenziale del sistema è l'omogeneità o meglio l'*univocità* degli indici. In altre parole, i rialzi di qualsivoglia indice devono sempre esprimere *miglioramento* della situazione economica: i ribassi, *peggioramento*. Perciò i fenomeni, il cui moto ascensionale è sintomo sfavorevole (emigrazione, fallimenti, disavanzo nel bilancio dello Stato, ecc.), vogliono essere interpretati, per via di proporzioni inverse o con altro artificio, da indici, non già crescenti, ma decrescenti. Invece

i fenomeni, il cui moto ascensionale è segno di sana economia e di sviluppo fisiologico, non patologico, del paese, si interpreteranno per via di proporzioni dirette con indici pure crescenti (esempio, i consumi, il corso dei titoli pubblici, il movimento ferroviario, ecc.).

§ 4. Veniamo alle questioni speciali. Ad alcune, per brevità, accenniamo soltanto; di altre indichiamo la soluzione più plausibile.

a) Per ciascun fenomeno, che si utilizza nella formazione di indici, importa anzitutto una scelta accurata degli elementi di *relazione generica e specifica*. Il crescere o diminuire dei protesti cambiari acquista un significato preciso solo quando risulti *più che proporzionale* al numero delle cambiali create. Avrà significato sfavorevole per l'economia nazionale il ribasso dei prezzi all'esportazione, quando i prezzi all'importazione si mantengono elevati, ciò volendo dire che il paese vende a buon mercato e compera caro, molto più se v'è motivo di credere che il rinvilio dei prodotti d'esportazione sia forzato (dovuto, ad es., a chiusura di mercati esteri, a nuove concorrenze internazionali, ecc.) e non sia invece la naturale conseguenza di abbondanti raccolti, di riduzioni di costo per effetto di progressi tecnici od altro (1). Avrebbe invece significato neutro o favorevole se nel contempo il livello dei prezzi alla importazione fosse disceso in proporzione o più che in proporzione.

In generale, poichè il fenomeno, preso in assoluto, non ha significazione ben determinata, è d'uopo far concorrere alla formazione dell'indice gli elementi da cui quello più immediatamente dipende. Così l'ammontare delle successioni si porrà in rapporto colla mortalità e, più propriamente ancora, colla mortalità dei maggiorenni, essendo notorio che i morti in minore età assai di rado dànno luogo a trasmissioni ereditarie e queste sono per lo più di scarsa entità. Il movimento ferroviario delle merci o dei viaggiatori si ragguaglierà alla popolazione e alla lunghezza delle linee, ecc. A parità d'altre circostanze, sono da preferire per la formazione degli indici i rapporti aventi carattere di maggior stabilità, in quanto colla loro costanza dimostrano di essere l'espressione di condizioni fondamentali della vita economica. Vi è poi un altro motivo di preferenza. Poniamo di ragguagliare gli *effetti in sofferenza* presso una grande Banca al totale del portafoglio. L'ammontare di cotesti effetti può variare da poche decine di migliaia di lire in epoche di prosperità a parecchi milioni in epoche di crisi. Trasformando i rapporti in indici si avrebbero variazioni enormi; da un anno all'altro l'indice della situazione economica (o soggettivamente parlando, l'indice

(1) Il deprezzamento, concomitante di progressi tecnici e di raccolti favorevoli, suole accompagnarsi ad una maggiore attività di scambi; non così, quando è forzoso per ostacoli doganali od altro. Nel caso in esame bisognerebbe combinare gli indici dei prezzi e quelli delle quantità e confrontare poi la serie ottenuta per le esportazioni con quella ottenuta per le importazioni.

del benessere) potrebbe trovarsi decuplicato o ventuplicato ovvero ridotto a un decimo o a un ventesimo di quel ch'era prima; e variazioni siffatte neutralizzerebbero le variazioni piccole e contrarie, ma non meno importanti di altri indici. Per evitare l'inconveniente, sarà buon partito ragguagliare il portafoglio presunto sano al totale del portafoglio, con che si otterrà una serie più stabile, cioè a piccole variazioni.

b) Il fenomeno, del quale ci vorremmo servire come d'indicatore della situazione economica, può essere parte di un fatto più complesso, così da non acquistare un deciso significato, senza la conoscenza almeno approssimata degli altri fenomeni, che hanno con esso un vincolo di solidarietà. Un eccesso di importazioni di merci sulle esportazioni non è sintomo nè buono nè cattivo, finchè non si conoscano gli altri elementi del *dare ed avere* internazionale, come i noli guadagnati dalla bandiera nazionale e dalla forestiera, le spese dei forestieri viaggianti, le rimesse degli emigrati, i titoli collocati all'estero o riscattati, ecc. Se vi fosse, ad esempio, motivo di ritenere che l'eccesso delle importazioni fu favorito da spedienti di finanza, che tennero artificiosamente depresso il cambio (come può avvenire colle emissioni di nuovi titoli all'estero fatte per pagare interessi di debiti vecchi), il carattere patologico dell'indice non potrebbe revocarsi in dubbio. Se lo sbilancio commerciale è concomitante di alti prezzi all'importazione e di bassi prezzi all'esportazione, vi sarà pure presunzione di un peggioramento economico. La fiducia insomma che su una molteplicità di indici gli errori debbano in gran parte compensarsi, non dispensa da una attenta analisi, caso per caso, del valore dei singoli indici.

c) La tendenza dell'indice può essere ottimista o pessimista, secondo il tempo che fa. In tempi difficili inclinano all'ottimismo i dati dei redditi desunti dai ruoli della ricchezza mobile, il cui aumento non è che il frutto di più sapiente persecuzione fiscale; i bilanci preventivi degli enti pubblici per la innegabile inclinazione degli amministratori a ritenere passeggiera la crisi e a ritardare provvedimenti impopolari; le sofferenze dei portafogli bancari, che in circostanze normali si pubblicano come stanno, in tempo di crisi forse si attenuano per non impressionare sfavorevolmente il pubblico. Diventano pessimiste le statistiche di consumi colpiti da forti dazi o imposte di fabbricazione, per maggiori quantità contrabbandate; le donazioni e successioni per maggiori somme sottratte al fisco, ecc. Inclinano poi all'ottimismo o al pessimismo, secondo i casi, i dati di cui si va anno per anno migliorando il processo di accertamento.

d) Non di rado l'interpretazione di un fenomeno come sintomo fisiologico o patologico può essere resa dubbia dal concorso di cause non economiche, che si mescolano colle economiche. Occorre eliminare quel tanto di variazioni che sembra imputabile alle prime. Il movimento ferroviario appare cresciuto per fatto di dislocamenti di truppe o

di elezioni politiche o perchè il raccolto delle uve, scarso nell'Alta Italia, abbondante nella Bassa, ha moltiplicati i trasporti di uve lungo le reti? Il rigore del metodo esige l'eliminazione di cotesti elementi. Oppure la causa di quel maggior movimento, il quale si dovrebbe interpretare come sintomo favorevole, è essa stessa di carattere patologico: come se nell'anno avessero avuto luogo trasporti eccezionali di emigranti? In tale ipotesi, sarà ancor più necessaria l'eliminazione, e se l'eliminazione non è possibile, si dovrà non dimenticare nell'elenco degli indici quello della emigrazione, il quale varrà a correggere il significato apparentemente favorevole dell' indice del movimento ferroviario.

e) Un elemento perturbatore nella valutazione del significato degli indici prescelti è l'*adattamento* degli individui alle condizioni create da fatti antecedenti. Ad ogni modificazione dello stato economico corrisponde una modificazione della nostra condotta, che persiste più o meno a lungo anche quando son cessate le cause, da cui fu provocata. Molti di coloro, che durante una crisi economica s'indussero a rinunziare a certi consumi, non riprendono le vecchie abitudini se non con lentezza. Caratteristico è il caso del giuoco del lotto in Italia. È fuor di dubbio che il disagio, che afflisse il nostro paese dal 1888 al 1896, obbligò molti ad economizzare sulle spese voluttuarie, non esclusa quella del giuoco; molti altri dal giuoco si astennero del tutto. Per parte degli uni e degli altri la ripresa delle vecchie abitudini non si verifica che a poco a poco, quantunque a partire dal 1896 le condizioni del paese siano in generale assai migliorate.

f) Altra causa di complicazioni è l'*imitazione*; non difficile ad eliminarsi quando si palesa con continuità regolare; difficile quando procede a scatti, in forma, per dir così, epidemica. Del primo modo fa esempio in Italia il progredire del risparmio; del secondo sono esempi le discontinuità gravi e improvvise nelle serie dell'emigrazione e degli scioperi. In generale se si tratta di diffusione del costume da strati superiori a strati inferiori per ricchezza o di diffusione nel territorio o dell'una o dell'altra promiscuamente, non parrebbe cattiva norma quella di tradurre in indici le *radici quadrate* dei numeri della serie originaria — poichè le aree di diffusione crescono a un dipresso come i quadrati dei raggi — salva sempre l'ulteriore elaborazione dei dati così ottenuti con altri elementi di relazione generica o specifica.

g) L'ordine logico di successione dei diversi indici, che si vogliono comporre in uno, richiede caso per caso un'analisi diligente, diretta a stabilire per ogni fenomeno quali sono i suoi precedenti causali, quali i concomitanti, quali i conseguenti necessari. Il corso dei cambi può essere preannunziato dalla previsione d'un cattivo raccolto del frumento o dei bozzoli; può accompagnarsi ad un rialzo dello sconto, ad una diminuzione delle riserve bancarie, ecc., ed essere susseguito da una maggiore esportazione di merci o di titoli. È facile intuire che un'analisi

ben condotta eleverebbe l'indice plurimo a sistema razionale di *previsioni* economiche.

h) Ma la questione forse più importante da risolvere è quella del *peso* o *coefficiente d'importanza* da assegnarsi ai diversi indici. Il procedimento della media aritmetica, che in via provvisoria abbiamo applicato nell'esempio fatto sopra, presuppone che si attribuisca ad ogni fenomeno la stessa importanza, come sintomo, che ad ogni altro. Senonchè è facile intendere che in concreto le cose stanno diversamente. La diminuzione di $^1/_{10}$ nel consumo del sale sarebbe sintomo ben altrimenti importante della diminuzione di $^1/_{10}$ nel consumo della birra. Una variazione sfavorevole dell'1 % nel corso dei cambi ha un significato ben più grave che una variazione dell'1 o anche del 10 % nell'emigrazione. Occorre dunque attribuire un peso o coefficiente d'importanza ad ogni sintomo; con quali criteri?

Riteniamo che il criterio più plausibile sia quello di assegnare a ciascun fenomeno un peso o coefficiente d'importanza *in ragione inversa del suo grado di variabilità.* Succede dei fenomeni dell'organismo sociale alcunchè di analogo a ciò che si verifica nell'organismo fisico degli individui. Le parti più essenziali del corpo umano (testa, cuore, ecc.) sono sotto l'aspetto delle dimensioni le più costanti; le parti meno essenziali sono invece le più variabili. Tra il capo del gigante e quello del nano intercede di solito molto minor differenza che tra le braccia o le gambe dell'uno e quelle dell'altro. Lo stesso per l'organismo sociale. I fenomeni connessi con certe condizioni essenziali di vita di un popolo hanno un campo di variabilità ristretto, senza di che la vita del corpo sociale sarebbe, per dir così, intermittente. Spaziano invece ordinariamente in un campo più largo i fenomeni, che hanno ragion d'essere in condizioni di secondaria importanza per l'organismo sociale. Naturalmente non si tratta di categorie fisse, nè è escluso che un fenomeno passi da una categoria all'altra, nè si pretende che la norma suggerita sia applicabile a rigore anche ai fenomeni dipendenti più da cause naturali che dall'opera dell'uomo. Così la variabilità notevole della produzione granaria non scemerebbe il coefficiente d'importanza di questo indice in un paese dall'agricoltura primitiva ed esportatore quasi unicamente di grano; lo scemerebbe però nei paesi europei, le cui rapide comunicazioni col resto del mondo e la varietà dei prodotti di ricambio hanno per risultato di colmare facilmente le deficienze dei raccolti proprii. Noi dunque sosteniamo il criterio proposto, fatta riserva per casi speciali. In una rappresentazione grafica delle serie di indici, tale criterio si attuerebbe prendendo come unità di misura sull'asse delle ordinate un segmento, eguale per tutti gli indici, corrispondente alla variazione media delle singole serie.

Nè è da credere che questo spediente, riducendo alla stessa ampiezza *media* le variazioni dei diversi fenomeni, converta l'indice composto in una inutile ripetizione e somma di indici semplici equivalenti. Equi-

valenza ci può essere, ma nelle serie come tali, non nei singoli momenti loro. Perocchè le variazioni, anche ridotte alla stessa ampiezza media, non saranno necessariamente *sincrone* da serie a serie, alcuni sintomi manifestandosi più tardi di certi altri; nè si corrisponderanno esattamente per altri caratteri (altezza dell'onda, durata, ecc.), ogni indice avendo un suo proprio grado di sensibilità, elasticità, ecc., ed essendo in varia misura affetto da influenze accidentali non eliminabili nel corso dell'elaborazione. È conforme alla natura delle medie, che una stessa media possa risultare da termini assai diversi. Sicchè, anche dopo applicati i coefficienti d'importanza nel modo che fu detto, le singole variazioni non essendo identiche o, per dirla in linguaggio geometrico, *sovrapponibili* da serie a serie, ciascuna servirà da correttivo delle altre nella formazione dell'indice composto.

Molti altri quesiti speciali si affacciano, che noi appena tocchiamo di volo: se e come si debba tener conto del diverso grado di certezza statistica dei fenomeni scelti per la formazione degli indici; se convenga col mezzo di interpolazioni parziali liberare anzitutto le serie dalle perturbazioni dovute a cause specialissime; come debbano entrare in calcolo i fenomeni, direm così, localizzati, sia perchè non si estendono geograficamente all'intero paese, sia perchè si producono nell'àmbito di classi sociali determinate; e così via.

La molteplicità e talora la difficoltà dei quesiti non deve far concludere affrettatamente per l'impossibilità del sistema dell'indice plurimo, nè indurre il lettore a pretenderne la soluzione in un'opera di carattere generale (1).

§ 5. Una maniera elegante di paragonare tra loro ed eventualmente comporre in uno gli indici del movimento economico sarebbe quella di calcolare le equazioni abbastanza approssimate delle singole serie, fissata una comune origine dei tempi, e restringere il confronto alle *parti variabili regolari* di esse serie, poichè le parti variabili irregolari si debbono presumere eliminate dal processo interpolatorio. Veggasi in proposito quel che dicemmo dei *Rapporti tra serie* (pag. 206). Fatta eguale a 100 la media aritmetica, si ridurrebbero in proporzione le costanti delle equazioni residue. L'identità di segno delle costanti che si corrispondono da serie a serie indicherebbe l'andamento parallelo;

(1) Men che meno vorrà il lettore giudicare dell'impossibilità del sistema da un informe abbozzo da me tentato prima per gli indici del movimento economico in Italia (*Giornale degli Economisti*, febbraio 1892), sul quale esercitò il suo spirito critico F. Coletti (*Rassegna di scienze sociali e politiche*, Firenze, luglio 1892) e da un altro abbozzo, un po' meno rudimentale, ideato per la provincia di Bari (*Dati statistici sul movimento economico in provincia di Bari, con totalizzatore e tavola grafica*. Bari, Avellino e C., 1892). Chi scrive non ha rinunziato all'idea di una terza prova da tentarsi con maggior rigore di metodo e con più ricco materiale statistico.

la opposizione di segno, l'andamento contrario. Del pari la grandezza relativa delle costanti starebbe a denotare la profondità od altezza relativa dell'onda corrispondente. Il quadro delle serie si riduce così in minimo spazio ad una tabella di costanti. Oltre a ciò converrebbe istituire un esame a parte dei residui eliminati nel processo interpolatorio, ossia delle parti presunte variabili irregolari delle serie, potendo darsi anche per queste delle identità di segno e delle correlazioni degne di nota.

§ 6. Senza entrare in più diffusi particolari sulla elaborazione degli indici singoli, concludiamo che il sistema dell'*indice plurimo* o *composto*, quando fosse razionalmente attuato, sottrarrebbe al variabile apprezzamento individuale alcuni fatti complessi e alle troppo facili generalizzazioni sostituirebbe una misura oggettiva e non arbitraria. Oltre alla comparabilità delle situazioni da tempo a tempo e da luogo a luogo, si avrebbe il vantaggio di conoscere in che rapporto sta ogni singolo fenomeno col gruppo degli affini o col concerto degli elementi, da cui si è ricavato l'indice composto. Sensibilità, elasticità, ordine di propagazione delle variazioni, distanza di tempo tra causa ed effetto, tendenza ottimista o pessimista, coefficienti d'importanza, ecc., sono aspetti che ben meritano l'analisi statistica, la quale può offrire all'economista nuove leggi empiriche o la verifica di leggi razionali già stabilite per via deduttiva. Ma per arrivare a tanto, il sistema dell'indice plurimo presuppone un organismo di vere e proprie monografie, in cui siano discussi indice per indice quei singoli punti. Secondo un'idea pratica, ch'io ho raccolto dalla privata conversazione col prof. M. Pantaleoni (1), converrebbe gettare le basi d'un buon *dizionario semiologico*, limitato per ora agli indici del movimento economico (poichè è chiaro che il sistema, *mutatis mutandis*, potrebbe immaginarsi applicato agli indici del movimento intellettuale e morale di un paese), nel quale si trovassero per ogni voce indicati: la fonte dei dati, il grado di certezza statistica, gli elementi di relazione generica e specifica, i fenomeni precursori, concomitanti e susseguenti, il campo di variabilità, la legge della serie per il periodo in osservazione, le probabilità di variazioni in un futuro assai prossimo ed altre note caratteristiche.

---

(1) Del PANTALEONI convien ricordare le osservazioni di Semiotica generale contenute nella prima parte dello scritto: *Sull'ammontare probabile della ricchezza privata in Italia*, Roma 1884, e quelle sui tre sistemi di semiologia economica, nella *Revue d'Économie politique*, ottobre 1892 (*Observations sur la sémiologie économique*). Veggasi altresì F. COLETTI, *Dell'Indice unico* (*Giornale degli Economisti*, febbraio-marzo 1903).

CAPO SESTO

Casi particolari di tabelle derivate.

TAVOLE DI MORTALITÀ E DI SOPRAVVIVENZA.

Sommario: § 1. Preliminari. — § 2. Metodo di Halley. — § 3. Metodo dei *censiti*. — § 4. Metodo di Hermann. — § 5. Metodo di Quételet. — § 6. Tavola di sopravvivenza della popolazione italiana. — §§ 7 e 8. Formole della vita media per una testa e per due teste. — § 9. Tavole di *nuzialità*, di *fecondità*, ecc.

§ 1. Si dicono *tavole di mortalità* i prospetti che indicano quanti individui su 100 o su 1000, che fanno il loro ingresso in un dato anno di età, muoiono nel corso dell'anno. Si dicono *tavole di sopravvivenza* quelle che, partendo da un numero fisso di nati nello stesso momento (ad esempio 100.000) indicano quanti di costoro sopravvivono all'età di 1 anno preciso, quanti all'età di 2 anni, quanti a quella di 3, ecc., fino ad estinzione del gruppo considerato.

Sia, ad esempio, il prospetto:

Età		Individui	
0 anni	. . . . . .	100.000	(nati)
1 anno	. . . . . .	79.570	(sopravviventi)
2 anni	. . . . . .	71.020	»
3 »	. . . . . .	67.240	»
4 »	. . . . . .	65.030	»
...		...	

Questo sarebbe un frammento di tavola di sopravvivenza.

S'intende subito che dalla tavola di sopravvivenza si può ricavarne una di mortalità, come da questa si potrebbe ricavare quella. Se per 100.000 nati abbiamo 79.570 superstiti all'età di un anno preciso, vuol dire che nel corso del primo anno sono morti 20.430 bambini, che ragguagliati ai 100.000 nati da cui provengono, dànno il quoziente di mortalità del 204,3 $^0/_{00}$. Così la differenza tra 79.570 (superstiti ad un anno d'età) e 71.020 (superstiti a 2 anni), differenza che è di 8550, corrisponde ai bambini morti nel corso del loro secondo anno d'età; e questi ragguagliati ai 79.570 che compirono il primo anno di loro età e fecero ingresso nel secondo, dànno il quoziente 107,4 $^0/_{00}$. E così di seguito. Sicchè avremmo la serie dei quozienti:

Età	Quozienti di mortalità
Da 0 a 1 anno	204,3 per mille
» 1 a 2 anni	107,4 »
» 2 a 3 »	53,2 »
» 3 a 4 »	32,8 »
...	...

E questa sarebbe una tavola di mortalità.

I metodi per la costruzione di queste tavole diversificano più o meno tra loro, secondo che lo statistico ha a propria disposizione solo una classificazione dei morti per età (*lista mortuaria*) o solo una classificazione per età dei viventi (*censimento*), ovvero ha ad un tempo l'una e l'altra, eventualmente accompagnate dai rilievi delle nascite, della immigrazione e dell'emigrazione. L'ideale sarebbe naturalmente quello di seguire una generazione di nati sino alla morte dell'ultimo superstite, registrando nominativamente i decessi nel primo anno di età, nel secondo, nel terzo, ecc. Ma, oltre che bisognerebbe aspettare un secolo per avere completa la tavola, grandissima sarebbe la difficoltà di seguire, individuo per individuo, le vicende di una intera generazione, senza perdere le traccie di nessuno fra i cambiamenti di residenza all'interno, l'emigrazione e le scomparse che rimangono ignorate. Dobbiamo dunque contentarci di metodi approssimativi, ma più spediti, e che del resto dànno in pratica risultati attendibili.

§ **2.** *Metodo di Halley.* — Supponiamo di avere a nostra disposizione null'altro che una semplice classificazione dei morti per età. Ignoto quindi il numero dei viventi nelle diverse classi di età; ignoto il numero delle nascite annuali, ecc. Come mai si determineranno i quozienti di mortalità, cioè le proporzioni dei morti agli esposti a morire nelle singole classi di età, se ignoriamo uno dei termini del rapporto, cioè il numero dei viventi?

L'astronomo Halley, al quale si deve la prima tavola calcolata sulle sole liste mortuarie di Breslavia (1693), partì dall'ipotesi di una *popolazione stazionaria*, nella quale pertanto il numero medio delle nascite bilancia quello delle morti, astrazion fatta dai movimenti migratori. Ammessa tale ipotesi, il totale *accertato* dei *morti* fornirà il numero *congetturale* dei *nati*. Se da questo numero congetturale di nati togliamo quello dei bambini morti in età da 0 ad 1 anno, il residuo ci darà il numero approssimativo dei *superstiti ad* 1 *anno preciso di età*. Se da questi superstiti ad un anno preciso d'età togliamo ancora i bambini decessi nel corso del loro secondo anno di vita, il nuovo residuo ci darà, sempre in via d'approssimazione, i *superstiti a* 2 *anni precisi di età*. Sottraendo successivamente quelli morti in età da 2 a 3 anni, otterremo i *superstiti a* 3 *anni*. E così di seguito. Si riesce per tal modo a costruire una tavola di viventi alle diverse età senza aiuto di censi-

mento e con una approssimazione al vero tanto maggiore quanto più l'ipotesi di una popolazione stazionaria corrisponde alla realtà.

Ma per una popolazione in via di sensibile progresso o regresso il procedimento indicato non reggerebbe più. Poniamo che essa sia da lunga data crescente e che ogni nuova generazione sia d'ordinario più numerosa di quella dell'anno antecedente; il metodo di Halley farebbe apparire più grande del vero la mortalità relativa delle età giovani, men grande del vero quella delle età avanzate. Perocchè nel totale dei decessi i bambini e gli adolescenti entrano in una proporzione ingrandita pel fatto stesso che essi provengono da schiere di nati recenti e più numerose di quelle da cui provengono gli adulti e i vecchi; e viceversa questi entrano nel totale dei decessi in una proporzione tanto più impicciolita, quanto men numerose in confronto delle attuali furono le lontane generazioni di nati da cui provengono.

In Italia, nel periodo 1890-1895, si ebbe una media annuale di 788.486 morti. Se a quell'epoca non avessimo conosciuto per fonte diretta quante erano le nascite annuali, partendo dall'ipotesi della popolazione stazionaria noi le avremmo supposte eguali a 788.486, cioè al totale delle morti. Da questo numero congetturale di nati tolgansi i 206.910 bambini morti — in media annuale durante il periodo in esame — tra la nascita e il compimento del primo anno d'età; il residuo 581.576 indicherà sempre in via congetturale i superstiti a un anno preciso d'età. Da questo residuo tolgansi ancora 82.815 bambini che risultarono morti da 1 a 2 anni; avremo 498.761 sopravviventi a 2 anni precisi. E così via.

I quozienti di mortalità si otterranno dividendo il numero dei morti d'ogni classe d'età per quello dei sopravviventi da cui derivano. Quindi:

Età	Nati e sopravviventi	Differenze ossia *morti*	Quozienti di mortalità
0 anni	788.486		
		206.910	262,4 per mille
1 anno	581.576		
		82.815	142,4 »
2 anni	498.761		
		36.304	72,3 »
3 »	462.457		
...	...	...	...
...	...	...	...
...	...	...	...

Senonchè in quel turno di tempo e anche per molt'anni addietro *le nascite effettive superarono il milione;* quindi è facile intendere che nel rapporto tra i morti infanti e i nati e in quello tra i morti nel corso del secondo, terzo, ecc., anno di età e i superstiti da cui provengono, il denominatore viene a trovarsi ingrandito; epperò i quozienti reali di mortalità debbono risultare attenuati in confronto di quelli ottenuti col metodo di Halley.

Aggiungasi che con questo metodo non si tien conto dell'influenza delle migrazioni; difetto tuttavia comune ad altri metodi.

§ **3.** *Metodo dei censiti.* — Abbiasi ora null'altro che un censimento della popolazione classificata per età; quindi nessun rilievo statistico delle nascite e delle morti annuali. Come costruire una tavola di mortalità?

Bisogna ancor supporre stazionaria la popolazione e ammettere che la differenza numerica tra i censiti di una data età e quelli dell'età successiva dipenda unicamente dalle morti avvenute nell'intervallo. Il censimento si fa servire così da tavola di sopravvivenza. Prendendo, ad esempio, alcuni dati dell'ultimo censimento italiano, il numero dei morti e i quozienti di mortalità risulterebbero così:

Età (anni)	Censiti	Morti (*calcolati per differenza*)	Età (anni)	Quozienti di mortalità
0-1	943.216			
		108.939	da $^1/_2$ a 1 $^1/_2$	115,5 per mille
1-2	834.307			
		41.004	» 1 $^1/_2$ a 2 $^1/_2$	49,1 »
2-3	793.303			
		10.922	» 2 $^1/_2$ a 3 $^1/_2$	13,8 »
3-4	782.381			
		19.107	» 3 $^1/_2$ a 4 $^1/_2$	24,4 »
4-5	763.274			
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...

Se la popolazione non è stazionaria, ma da lunga data in progresso, avviene che i censiti di una data età sono men numerosi di quelli dell'età immediatamente anteriore, non solo a causa delle morti avvenute nell'intervallo, ma altresì perchè la classe più vecchia proviene di solito da una generazione di nati men numerosa di quella da cui proviene la classe più giovane. Il metodo anzidetto, considerando adunque la differenza numerica tra l'una e l'altra classe come se fosse unicamente dovuta a mortalità, esagera il numeratore della frazione, colla quale si esprime il rapporto tra morti ed esposti a morire, e perciò dà quozienti più elevati del vero.

Potrebbe darsi che i bambini, poniamo, fra i 4 e i 5 anni d'età, venuti al mondo in un'annata molto feconda, si trovassero in numero quasi eguale o anche superiore di quelli da 3 a 4 anni, nati in un anno di scarsa fecondità; e allora le cose starebbero come se non si fosse verificata mortalità alcuna nell'intervallo dai 3 $^1/_2$ ai 4 $^1/_2$ anni o come se un certo numero di morti fosse ritornato in vita! A tale inconveniente si rimedia col formare classi d'età più comprensive (ad esempio, quinquennali), rendendo così più probabile la compensazione di annate feconde e poco feconde.

§ **4.** *Metodo del confronto tra le nascite e le morti.* — Abbiansi a disposizione i dati dei morti classificati per età e quelli delle nascite. Facciamo, ad esempio, la semisomma dei nati nel 1896 e nel 1897; potremo riguardare questo numero come quello di altrettanti bambini nati contemporaneamente al 1° gennaio 1897. A tal numero ragguaglieremo quello dei bambini morti in età da 0 a 1 anno nel corso del 1897. Il *rapporto* tra i due numeri ci darà il *quoziente di mortalità* nel primo anno d'età; la *sottrazione* del secondo dal primo ci darà i *superstiti* a 1 anno preciso d'età, al 1° gennaio 1898. Da questi superstiti sottrarremo i bambini morti in età da 1 a 2 anni nel corso del 1898: il residuo indicherà i *superstiti* al 1° gennaio 1899 in età di due anni precisi. Da questi ancora toglieremo i morti in età da 2 a 3 anni nel corso del 1899 e avremo i superstiti in età di 3 anni al 1° gennaio 1900 e così di seguito. Ecco un prospetto desunto dal *Movimento della popolazione secondo gli atti di stato civile in Italia per il* 1902.

Anni	Semisomma delle nascite di 2 anni consecutivi	Morti in età da 0 a 1 anno	Superstiti ad 1 anno d'età	Morti in età da 1 a 2 anni	Superstiti a 2 anni d'età.	Morti in età da 2 a 3 anni	Superstiti a 3 anni d'età	Morti in età da 3 a 4 anni
1896	1.098.677	—	—	—	—	—	—	—
1897	—	180.804	—	—	—	—	—	—
	1.085.961	—	917.873	—	—	—	—	—
1898	—	183.460	—	73.092	—	—	—	—
	1.079.316	—	902.501	—	844.781	—	—	—
1899	—	168.244	—	66.018	—	28.807	—	—
	1.077.967	—	911.072	—	836.483	—	815.974	—
1900	—	185.823	—	73.218	—	31.649	—	16.803
	1.062.569	—	892.144	—	837.854	—	804.834	—
1901	—	175.855	—	64.828	—	27.092	—	14.451
	1.075.419	—	886.714	—	827.316	—	810.762	—
1902	—	187.816	—	69.767	—	29.972	—	15.905

I quozienti di mortalità risultano quindi:

Anni	Dalla nascita a 1 anno	Da 1 a 2 anni	Da 2 a 3 anni	Da 3 a 4 anni
1897	164,57 per mille	—	—	—
1898	168,94 »	79,63 p. mille	—	—
1899	155,88 »	73,15 »	34,10 p. mille	—
1900	172,38 »	80,36 »	37,84 »	20,59 p. mille
1901	165,50 »	72,67 »	32,33 »	17,96 »
1902	174,64 »	78,68 »	36,23 »	19,62 »

Limitato alle prime classi d'età, questo metodo (detto di *Hermann*) offre il vantaggio di calcolare i sopravviventi cogli elementi forniti dai registri dello stato civile, che sono senza confronto più precisi di quelli forniti dal censimento. Inoltre nei limiti delle prime età i movimenti migratori sono poco rilevanti, figurando in essi per lo più gli individui adulti, atti al lavoro, in maggioranza celibi, che partono soli. Quindi, anche per tale rispetto, il metodo in esame è consigliabile.

§ **5.** *Metodo di Quételet* (o della *decima*). — Presuppone un buon censimento dei viventi e una regolare registrazione dei morti, gli uni e gli altri classificati per età. Ciò posto, si ragguagliano i morti d'ogni età, non ai viventi della stessa classe d'età, che in qualsiasi istante sono una cifra al netto delle perdite subìte, ma agli *esposti a morire*, che corrispondono in via approssimativa ai viventi aumentati della metà dei morti coetanei (1). Veggasi in proposito quanto abbiamo scritto a pag. 117. Dalla tavola di mortalità così ottenuta si passa poi facilmente alla tavola di sopravvivenza.

§ **6.** *Tavole di sopravvivenza della popolazione italiana.* — In base al censimento del 1881 e alla mortalità media del dodicennio 1876-1887 fu, a cura della Direzione generale della Statistica in Roma, costruita una prima e completa tavola di sopravvivenza della popolazione italiana. Per i primi 5 anni di età si è fatto il calcolo dei sopravviventi col metodo di Hermann, il censimento del 1881 avendo mostrato troppo gravi lacune nel rilievo dei bambini. Dai 5 anni ai 60 il metodo della decima venne applicato nella sua forma men corretta, cioè col ragguaglio dei morti d'ogni età ai censiti della classe corrispondente, anzichè agli esposti a morire; la qual cosa ha prodotto una serie di piccoli errori che, cumulandosi, fecero risultare più ridotte, che non avrebbero dovuto essere, le schiere dei superstiti. Dai 60 anni in poi il metodo della decima fu applicato nella sua forma corretta. Inoltre, essendosi proceduto dai 6 anni in poi per gruppi quinquennali, i quozienti di mortalità medii per ogni gruppo si attribuirono all'età di mezzo, e i quozienti di mortalità per le singole altre età, anno per anno, vennero calcolati per interpolazione parabolica di secondo grado dai 6 ai 13 anni, lineare dai 13 agli 88. Infine al di là del 90° anno, non potendosi utilizzare i dati greggi dei censiti e dei morti, che avrebbero dato il risultato inverosimile di una mortalità relativa più mite, anzichè più

---

(1) Questa correzione fu suggerita dal Barone di Wrede nelle formole adottate dalla Direzione di Statistica di Svezia fino dal 1851 e sostenuta dal De Baumhauer (Olanda), dal Lund (Danimarca) e dal Bodio (Italia). Però, data l'assai ineguale distribuzione dei morti da 0 a 1 anno nel corso dei primi dodici mesi d'età, la correzione indicata non basta per questa classe; conviene ricorrere daccapo al metodo di Hermann.

grave di quella del gruppo da 85 a 90 anni, si suppose che i quozienti di mortalità continuassero a crescere nella stessa ragion aritmetica, secondo la quale avevan progredito nell'intervallo tra gli 80 e i 90 anni.

La nuova tavola di sopravvivenza e di mortalità della popolazione italiana, costruita in base al censimento 9-10 febbraio 1901 e alle medie annue dei morti nel periodo 1899-1902, si distingue dalla precedente per qualche maggior rigore nel metodo. Con opportuni calcoli la classificazione per età dei censiti venne riportata al 1° gennaio 1901 e si procedette per il resto nella maniera già praticata per la tavola anteriore, salvo che il metodo corretto del confronto tra i morti e gli esposti a morire si applicò a partire dal 50°, anzichè dal 60° anno d'età e per il calcolo dei quozienti anno per anno, dai 15 ai 90 (da inserire fra i quozienti corrispondenti all'età centrale dei gruppi originari, che comprendevano or tre, or quattro, or cinque classi annuali), si adottò l'interpolazione parabolica di terzo grado. Sarebbe desiderabile, in una nuova edizione della tavola, che il metodo corretto del confronto tra morti ed esposti a morire, venisse esteso anche alle età fra 6 e 50 anni.

Diamo alla pagina seguente la tavola di sopravvivenza e di mortalità per la popolazione italiana senza distinzione di sesso, accompagnandola coi dati della vita probabile e della vita media (1).

---

(1) V. *Movimento della popolazione secondo gli atti dello stato civile nell'anno* 1902 (Direz. gen. della Statistica, Roma 1904). Introduzione, pag. LIX e seguenti.

TAVOLA DI SOPRAVVIVENZA E DI MORTALITÀ PER LA POPOLAZIONE ITALIANA.

Età	Sopravviventi	Quozienti di mortalità ‰	Vita probabile in anni e mesi	Vita media in anni e mesi	Età	Sopravviventi	Quozienti di mortalità ‰	Vita probabile in anni e mesi	Vita media in anni e mesi
0	100.000	167,17	52,11	43	51	51.393	13,36	20,7	20,1
1	83.283	76,54	60,10	50,4	52	50.706	14,18	19,9	19,4
2	76.908	35,87	62,2	53,7	53	49.987	14,44	18,11	18,7
3	74.149	19,61	62,2	54,8	54	49.265	15,03	18,2	17,10
4	72.695	13,58	61,7	54,10	55	48.524	16,04	17,8	17,2
5	71.711	9,01	60,11	54,6	56	47.746	17,48	16,6	16,5
6	71.065	7,56	60,1	54	57	46.914	19,12	15,9	15,8
7	70.528	5,78	59,4	53,5	58	46.017	21,30	15	15
8	70.120	4,68	58,5	52,9	59	45.037	23,71	14,9	14,4
9	69.792	4,00	57,5	51,10	60	43.969	26,68	13,5	13,7
10	69.512	3,68	56,8	51,1	61	42.796	28,91	12,6	13
11	69.256	3,27	55,8	50,4	62	41.559	31,66	12	12,4
12	69.030	3,30	54,10	49,6	63	40.243	34,51	11,4	11,10
13	68.802	3,25	53,11	48,8	64	38.854	37,52	10,7	11,3
14	68.578	3,76	52,11	47,10	65	37.396	40,83	10	10,7
15	68.320	4,10	52	47	66	35.869	44,38	9,5	10
16	68.040	4,46	51,1	46,1	67	34.277	48,29	8,10	9,5
17	67.736	4,95	50,2	45,5	68	32.622	53,56	8,3	8,11
18	67.401	5,41	49,3	44,7	69	30.875	59,78	7,9	8,5
19	67.036	5,88	48,5	43,10	70	29.031	67,12	7,1	7,10
20	66.642	6,25	47,6	43,1	71	27.082	74,57	6,6	7,5
21	66.225	6,70	46,8	42,4	72	25.064	82,80	6,3	6,11
22	65.781	6,98	45,10	41,8	73	22.987	90,01	5,10	6,6
23	65.322	7,04	44,11	40,11	74	20.918	98,10	5,5	6,3
24	64.862	7,08	44	40,2	75	18.866	107,12	5	5,10
25	64.403	7,12	43,2	39,7	76	16.844	116,98	4,7	5,5
26	63.944	7,14	42,4	38,10	77	14.873	127,74	4,3	5,1
27	63.487	7,15	41,5	38,1	78	12.973	139,26	4	4,10
28	63.033	7,15	40,7	37,3	79	11.166	151,72	3,6	4,6
29	62.582	7,17	39,8	36,6	80	9.472	164,98	3,5	4,2
30	62.133	7,20	38,10	35,10	81	7.909	179,12	3,2	3,11
31	61.685	7,24	38	35,2	82	6.492	194,18	3	3,9
32	61.238	7,32	37,1	34,5	83	5.232	208,37	2,11	3,7
33	60.790	7,38	36,2	33,7	84	4.142	223,92	2,9	3,5
34	60.341	7.47	35,4	32,9	85	3.214	239,06	2,5	3,2
35	59.890	7,60	34,5	32,2	86	2.445	254,37	2,2	3
36	59.435	7,78	33,6	31,4	87	1.823	269,80	2,1	2.10
37	58.973	8,04	32,8	30,8	88	1.331	285,22	1,11	2,8
38	58.499	8,35	31,10	29,10	89	951	300,63	1,10	2.7
39	58 010	8,61	30,11	29,1	90	665	310,53	1,10	2,6
40	57.510	8,88	30	28,4	91	459	317,90	1,9	2,5
41	57.002	9,02	29,2	27,6	92	313	326,41	1,8	2,4
42	56.488	9.20	28,4	26,10	93	211	333,06	1,6	2,3
43	55.968	9,46	27,5	26,1	94	140	341,60	1,7	2,1
44	55.438	9,64	26,7	25,4	95	92	348,20	1,7	2
45	54.903	9,90	25,9	24,7	96	60	355,61	1,6	1,11
46	54.359	10,17	24,11	23,9	97	38	362,54	1,6	1,8
47	53.806	10,50	24	23,1	98	24	370,88	1,4	1,8
48	53.241	11,05	23,2	22,3	99	15	382,80	1,1	1,1
49	52.652	11,67	22,8	21,7	100	9	—	0,6	0,6
50	52.037	12,37	21,5	20,8	—	—	—	—	—

§ 7. Per *vita probabile* s'intende il periodo di tempo che deve trascorrere perchè gli individui di una medesima classe di età siano ridotti a metà per successive morti. Così i 100.000 nati si trovano ridotti a 49.987 individui, cioè quasi giusto a metà, in capo a 53 anni; la vita probabile alla nascita si dirà dunque di 53 anni (più precisamente, secondo la tavola, 52 anni e 11 mesi). A 31 d'età la vita probabile è di 38 anni; infatti nella tavola i 61.685 individui trentunenni si trovano ridotti a 30,875, cioè quasi giusto alla metà, al raggiungimento del 69° anno d'età (69 — 31 = 38).

La vita probabile raggiunge il suo massimo nel terzo anno d'età (essa è allora di 62 anni e due mesi), e ciò perchè i bambini, i quali riuscirono a superare i primi e più gravi pericoli dell'infanzia, hanno davanti a sè un lungo periodo di mite mortalità. Dopo il quarto anno la vita probabile diminuisce lentamente dapprima, rapidamente poi; man mano che i superstiti hanno dinnanzi a sè un periodo più breve di mite mortalità e vanno incontro alle alte quote della vecchiaia.

Diverso è il concetto della *vita media*. Per « vita media » s'intende il numero d'anni che ancor vivrebbe ogni individuo di una data classe d'età, se tutti i componenti di questa mettessero insieme gli anni che loro restano a vivere fino ad estinzione del gruppo e se li ripartissero in parti eguali. In altri termini, la vita media si ottiene moltiplicando il numero dei morti di ogni successiva classe d'età per il numero d'anni vissuti dopo l'età che si è presa come punto di partenza, sommando tutti questi prodotti e dividendo il risultato per il numero dei viventi della classe da cui si è partiti.

Sia V_x la classe iniziale, di cui si vuol calcolare la vita media; la differenza tra essa e la classe successiva V_{x+1} sarà costituita dai morti nel corso dell'anno, i quali saranno stati presenti e vivi in detto anno all'incirca per 6 mesi in media (= $1/2$ anno). La differenza tra la classe di superstiti V_{x+1} e la V_{x+2} sarà costituita dai morti nel corso del secondo anno, i quali avranno vissuto, a partire dall'età x, un anno e mezzo (= $3/2$ di anno). La differenza tra la classe di superstiti V_{x+2} e la V_{x+3} sarà costituita dai morti nel corso del terzo anno, i quali avranno vissuto a partire dall'età x due anni e mezzo (= $5/2$ di anno). E così via.

Per guisa che tutto il gruppo considerato vivrà in totale anni:

$$\frac{1}{2}(V_x - V_{x+1}) + \frac{3}{2}(V_{x+1} - V_{x+2}) + \frac{5}{2}(V_{x+2} - V_{x+3}) + \dots$$

ossia, risolvendo le parentesi e semplificando, anni:

$$\frac{1}{2} V_x + V_{x+1} + V_{x+2} + V_{x+3} + \dots + V_{x+n}$$

essendo $x + n$ l'estremo anno d'età, al quale si trovano ancora dei superstiti secondo la tavola di sopravvivenza. Di solito si ha:

$$x + n = 100.$$

Quindi in media ogni individuo della classe V_x avrà vissuto, a partire dall'età x, anni:

$$\frac{\frac{1}{2} V_x + V_{x+1} + V_{x+2} + \ldots + V_{x+n}}{V_x}$$

formola che può anche scriversi così:

$$\frac{V_{x+1} + V_{x+2} + \ldots + V_{x+n}}{V_x} + \frac{1}{2}.$$

La vita media non coincide colla vita probabile; la supera nelle età avanzate, le è inferiore nelle età giovani e medie (1). Secondo la tavola italiana, il momento in cui l'una e l'altra si equivalgono è il 58° anno di età. La cosa del resto si capisce facilmente. Siano, ad esempio, degli individui settantenni, la cui vita probabile è di 7 anni. Ciò vuol dire che una metà di essi scompare nel settennio fra i 70 e i 77; mentre l'altra metà richiede per estinguersi un periodo più lungo, 23 anni almeno, se l'ultimo superstite arriva ai 100 anni. La vita media qui, per quanto la rapidità di esaurimento del gruppo cresca coll'età, deve riuscire alquanto più lunga della vita probabile. Per le età giovani, l'opposto. Adolescenti quindicenni si trovano ridotti a metà al 67° anno; la prima metà di essi impiega dunque per estinguersi 52 anni e tanta è la lor vita probabile. La seconda metà ne impiegherebbe soli 33, quanti corrono fra 67 e 100. Qui dunque la vita media deve risultare più breve della vita probabile.

Un concetto da tenersi distinto dagli anzidetti è quello dell'*età media dei morti* di un medesimo anno, i quali provengono da generazioni diverse di nati, tanto meno numerose quanto più remote, se la popolazione è da lunga data crescente; tanto più numerose quanto più remote, se la popolazione è da lunga data decrescente. È ovvio che vita media calcolata alla nascita ed età media dei morti non coincidono se non nell'ipotesi di una popolazione stazionaria. Se questa è in progresso, allora nel totale annuo dei morti i bambini entrano in una proporzione più considerevole di quella che risulterebbe da una tavola di sopravvivenza e colla loro bassa età deprimono la media generale.

§ 8. Per gli scopi pratici delle *assicurazioni*, *pensioni*, ecc., conviene anche il calcolo della *vita media per due teste*, cioè della durata

(1) La coincidenza della vita probabile e della vita media si avrebbe solo nell'ipotesi che il numero dei sopravviventi scemasse di eguali quantità col progredire dell'età d'anno in anno. Nemmeno può dirsi che la vita media stia in un rapporto costante colla vita probabile. Questo avverrebbe solo nel caso che la serie dei sopravviventi fosse formata di termini decrescenti in ragion geometrica (Veggasi *Movimento dello stato civile*, anno 1887, pag. LXXII, nota).

media di convivenza per due persone aventi la stessa età o età diversa (marito e moglie, padre e figlio, ecc.). La formola da applicarsi è la seguente:

$$D_{x,\,y} = \frac{V_{x+1}\,.\,V_{y+1} + V_{x+2}\,.\,V_{y+2} + \ldots + V_{x+n}\,.\,V_{y+n}}{V_x\,.\,V_y} + \frac{1}{2}$$

dove $D_{x,\,y}$ significa la durata della media convivenza di individui dell'età x con individui dell'età y; V_x, V_{x+1} ... sono i sopravviventi all'età x, $x+1$, ecc. della categoria a cui appartengono i primi individui; e V_y, V_{y+1} ... i sopravviventi alle età y, $y+1$, ecc. della categoria cui appartengono i secondi (potendoci essere, oltre che differenza d'età, anche differenza di sesso, di stato civile od altro).

Con questa formola il Perozzo ha calcolato la durata media dei matrimoni in Italia. Però il suo calcolo basato sulla vecchia tavola di sopravvivenza dei maschi e delle femmine in generale, senza distinzione di stato civile, diede risultati probabilmente inferiori al vero, essendo notorio che i coniugati hanno probabilità di vita alquanto superiori a quelle dei celibi e dei vedovi coetanei (1).

§ 9. Con procedimenti analoghi a quelli indicati per le tavole di mortalità e di sopravvivenza si sono pure costruite tavole di *nuzialità*, di *fecondità*, di *disoccupazione*, ecc. (2). Ma non ci pare il caso di entrare in particolari metodologici per argomenti aventi molta affinità con quello tipico, che fu ora trattato; come non crediamo opportuno discorrere delle applicazioni matematiche a questioni di assicurazioni e simili, che sono di competenza degli attuari.

---

(1) Veggasi la Nota di L. Perozzo sulla *Durata media dei matrimoni*, nel *Movimento dello stato civile in Italia nel* 1889, Appendice, pag. LIX e seguenti.

(2) Se ne possono vedere esempi nelle Statistiche municipali di Berlino (*Statistisches Jahrbuch der Stadt Berlin*), annate posteriori al 1893. In quelle dal 1894 al 1896 si trovano copiosi particolari sul metodo, col quale fu costruita la tavola di disoccupazione, da noi riprodotta nei nostri *Principii di Demografia*, pag. 191.

CAPO SETTIMO

Comparazione e critica dei dati elaborati.

Sommario: § 1. Artifici di comparazione. — § 2 e 3. Formazione di gruppi scelti ed esempi. — § 4. I confronti internazionali e l'opera dell'Istituto internazionale di Statistica.

§ 1. La elaborazione dei dati primitivi per mezzo dei procedimenti aritmetici, geometrici e algebrici spiana la via ad ulteriori còmpiti della comparazione e della critica.

Condizione prima per la comparabilità di elementi statistici è, come sappiamo, l'*omogeneità*. Ma questa può intendersi in tanti modi, che anche i fatti più diversi diventano sotto qualche aspetto paragonabili tra loro. A prima giunta si direbbe, ad esempio, che tra consumo del tabacco, suicidii e partecipazione degli elettori alle elezioni politiche, non ci sia comparazione immaginabile. Pure, se ponessimo il quesito così: « È più disuguale da provincia a provincia, in Italia, il consumo del tabacco, o la frequenza del suicidio, o la partecipazione degli elettori alle elezioni politiche? », la possibilità del paragone balzerebbe subito all'occhio. Nè sarebbe difficile la soluzione del quesito. Si dispongano le 69 provincie del regno, prima secondo la serie decrescente delle quote di consumo del tabacco, poi secondo quella decrescente delle quote di suicidii, poi ancora secondo quella decrescente delle quote di votanti; si descrivano in diagramma le tre serie (1) e s'interpolino tre rette. L'inclinazione maggiore o minore di queste rette dirà per quale dei fenomeni in esame vi ha maggiore o minor disformità o disuguaglianza da provincia a provincia. Ecco un modo di comparazione che potrebbe dar luogo a curiose graduatorie.

Per conseguire l'omogeneità convien ricorrere spesso ad artifizi, di cui difficilmente si potrebbe dare una teoria generale. Molto dipende dall'intuito dello statistico.

Il Galton, come abbiamo visto, per rendere paragonabili nei riguardi delle medie stature le figliuolanze diversamente assortite per sesso, ha in certo modo mascolinizzate le stature delle femmine, accrescendole di circa $^1/_{12}$. Per rendere paragonabili i consumi annuali di famiglie ope-

---

(1) S'intende, prendendo sull'asse delle ascisse segmenti successivi proporzionali alla popolazione delle singole provincie, e sull'asse delle ordinate un segmento comune per rappresentare le quote medie generali dei tre fenomeni considerati.

raie diversamente assortite per sesso ed età dei componenti, il Rubin adottò per le diverse età una scala proporzionale al numero di calorie prodotte in 24 ore colla nutrizione ordinaria; sicchè, ragguagliato a 100 il numero spettante al maschio adulto, quello del maschio a 16 anni risulterebbe di 76; a 12 anni di 60; a 10 anni di 50; a 5 anni di 43; a 2 anni di 30. Per le femmine, le stesse cifre ridotte di $^1/_5$ (1).

Il Körosi per rendere paragonabili le quote generali di mortalità delle popolazioni europee, le quali constano in diverse proporzioni di bambini, adulti e vecchi, ha suggerito giustamente di fissare una *standard population*, cioè una popolazione tipica per la proporzione delle diverse classi di età, alle quali classi si applicano i quozienti specifici di mortalità accertati nei diversi paesi, e dalla somma dei prodotti divisa per il totale della popolazione-tipo si ricavano i quozienti generali (2). Potremmo così ricordare molti altri spedienti del genere.

§ **2.** La formazione di *gruppi scelti* apre pure la via a svariati modi di comparazione. Limite certo alla formazione di tali gruppi non si può assegnare; l'essenziale è di non perdere di vista i più importanti caratteri d'opposizione tra gli elementi rilevati. In quanti modi non potremmo noi rimaneggiare i dati delle statistiche doganali? Merci rincarite, invariate di prezzo e deprezzate, da un anno o da un periodo all'altro — merci soggette ad alti dazi, a dazi miti e merci esenti — materie greggie, prodotti di mezza lavorazione, prodotti compiuti — prodotti di assoluta importazione, di assoluta esportazione e prodotti che sono, a vari gradi, più importati che esportati o viceversa — derrate i cui scambi hanno un massimo o un minimo di attività in certi mesi e derrate che l'hanno in certi altri mesi, e così via; ecco altrettanti gruppi scelti che si possono formare cogli elementi originari della rilevazione, a seconda delle questioni da risolvere.

Poichè siamo in argomento di statistiche commerciali, proponiamoci il problema: *quale sia stato l'aumento reale di protezione portato dalla tariffa italiana del* 1887 (*e dalle convenzioni basate su di essa*) *in confronto della tariffa più moderata del* 1878 (*e relative convenzioni*).

All'uopo formeremo per due periodi ben distinti, il 1881-1883 e il 1891-1893, i *gruppi scelti* seguenti:

1° Prodotti esenti di dazio all'importazione in entrambi i periodi;
2° Prodotti esenti nel primo periodo e tassati nel secondo;
3° Prodotti tassati nel primo ed esenti nel secondo;
4° Prodotti tassati in entrambi i periodi.

---

(1) Marcus Rubin, *Consommations de familles d'ouvriers danois* (*Bulletin de l'Institut international de Statistique*, Tome XIII, 3ème Livraison, Roma, anno 1903).

(2) Veggasi *Movimento della popolazione. Confronti internazionali*, parte II, pag. XXIII (Direzione generale della Statistica, Roma 1897).

PRODOTTI	Medie annue all'importazione (in milioni di lire)	
	1881-1883	1891-1893
Esenti in entrambi i periodi . . .	370,2	380,2
Esenti nel 1° e tassati nel 2° . . .	54,1	52,0
Tassati nel 1° ed esenti nel 2° . .	33,7	39,6
Tassati in entrambi i periodi . . .	748,0	621,3

Nonostante la discesa generale dei prezzi, che si è verificata nell'intervallo e che perturba il confronto, riesce visibile la contrazione degli scambi per quei prodotti, di cui furono aggravati i dazi o che cessarono di essere esenti (4° e 2° gruppo) e l'espansione per quelli che conservarono l'esenzione o furono sgravati di dazi (1° e 3° gruppo).
Ora limitando l'analisi al 4° gruppo e determinando le somme riscosse a titolo di dazio doganale sui prodotti che ne facevan parte, risulta che il rapporto del dazio al valore di detti prodotti è salito da 19,62 % a 34,47 % da un periodo all'altro. Infatti si ebbe:

Periodi	Dazio riscosso (in milioni di lire)	Valore delle merci importate (in milioni di lire)	Rapporto del dazio al valore
1881-1883 (media annua) . . .	146,8	748,0	**19,62 %**
1891-1893 (») . . .	214,2	621,3	**34,47 »**

Ma è pure importante sceverare nella larga categoria delle merci, che erano tassate all'importazione in entrambi i periodi, quelle soggette a dazi di natura propriamente *fiscale*, da quelle soggette a dazi *protettori*. Tra le prime poniamo il petrolio, i coloniali, le droghe e i tabacchi; tra le seconde i prodotti agricoli o industriali che entravano nel 4° gruppo. Otterremo così nuovi *gruppi scelti*.

Petrolio, coloniali, droghe e tabacchi.

Periodi	Dazio riscosso (in milioni di lire)	Valore all'importazione (in milioni di lire)	Rapporto del dazio al valore
1881-1883 (media annua) . . .	84,2	102,9	**81,82 %**
1891-1893 (») . . .	123,9	79,9	**155,15 »**

Altri prodotti agricoli o industriali tassati in entrambi i periodi.

Periodi	Dazio riscosso (in milioni di lire)	Valore all'importazione (in milioni di lire)	Rapporto del dazio al valore
1881-1883 (media annua) . . .	62,6	645,1	**9,70 %**
1891-1893 (») . . .	90,3	541,4	**16,67 »**

L'aumento *apparente* della protezione dovuto alla tariffa del 1887, più o meno temperata dai trattati, risulta eguale alla differenza tra 9,70 e 16,67 %, osservata limitatamente a quei prodotti, a cui riguardo il dazio di confine ha scopo proprio di difesa dalla concorrenza estera. E diciamo aumento *apparente*, perchè in parte esso si sarebbe verificato

in maniera automatica pel fatto stesso del deprezzamento generale subito dalle merci d'importazione nell'intervallo considerato. Se una merce è soggetta al dazio fisso di lire 40 al quintale, sia che essa valga 100 lire il quintale, sia che ne valga solo 80 o 60, è ovvio che ragguagliando il dazio al valore noi avremo, come conseguenza del deprezzamento, un rapporto crescente del primo al secondo; da 40 %, il rapporto salirà a 50 % e a 66,67 %. Orbene, chi tenga conto del rinvilio del 20 % verificatosi da un periodo all'altro nella generalità dei prodotti d'importazione, troverà che le quantità introdotte dal 1891 al 1893 avrebbero rappresentato ai prezzi medii del 1881-1883 un valore annuo di circa 677 milioni, anzichè di 541 soltanto; e in tal case il rapporto tra il dazio riscosso (90,3 milioni) e il valore importato discenderebbe da 16,67 % a 13,34 %. In altri termini, la differenza tra 9,70 % (misura media della protezione nel periodo 1881-1883) e la quota ora trovata di 13,34 %, differenza che si riduce a 3,64 %, starebbe a rappresentare l'aumento *reale* di protezione dovuto alla politica doganale inaugurata colla tariffa del 1887; e la differenza tra 16,67 e 13,34 % = 3,33 % starebbe invece a denotare l'aumento *automatico* del rapporto del dazio al valore, da imputarsi al ribasso generale dei prezzi.

Si potrebbe pensare che l'aumento di protezione risulti piccolo, solo perchè i prodotti più duramente trattati dai nuovi dazi vennero importati in quantità minori di prima, laddove quelli meno colpiti rappresentarono all'importazione nel secondo periodo una parte preponderante. L'obbiezione contiene certamente un po' di vero; ma solo un poco. Infatti (per citare una sola, ma importante voce del nostro commercio internazionale) il grano, il cui dazio passò da lire 1,40 nel 1881-1883 a lire 5,00 il quintale nel 1891-1893, fu importato durante questo secondo periodo in quantità, non minori, ma molto maggiori che nel primo. Ad ogni modo, per dirimere questo dubbio converrebbe ancor procedere col metodo dei *gruppi scelti*, il solo che possa dare risultati sicuri e che non incontra altro limite se non quello dei piccoli numeri, verso cui porta naturalmente la specializzazione dei dati.

§ **3.** I criteri per la formazione di gruppi scelti possono attingersi ad ogni ordine di conoscenze concrete o anche ad ipotesi non troppo disformi dalla realtà dei casi. Circa l'ufficio delle ipotesi in Statistica diremo più ampiamente in fine dell'opera; qui crediamo non inutile recare un esempio che è a proposito del tema.

Si tratti di confrontare il numero degli *impiegati addetti ai servizi civili* nel regno attuale con quello dei sette staterelli, in cui prima del 1860 era divisa l'Italia.

Un matematico ed economista insieme, cui fosse proposto il problema della riduzione di personale amministrativo, che a parità d'ogni altra circostanza dovrebbe aspettarsi dalla fusione di sette Stati in

uno, comincerebbe certo dal ridurre le ipotesi alla più semplice espressione. Immaginando due Stati di forma geometrica tipica, p. es., circolare, uniformi per territorio, distribuzione di popolazione, civiltà, ecc., ma l'uno sette volte più esteso e popolato dell'altro, egli distinguerebbe i servizi amministrativi che *caeteris paribus* crescono in ragione del crescere della periferia o linea di confine, da quelli che crescono invece in ragion di superficie (1) e da quelli ancora che crescono in ragion composta delle precedenti o in una ragione qualsiasi da determinarsi caso per caso. In breve: servizi *periferici*, servizi *diffusi* e servizi *centrali*.

Evidentemente rientrano nella prima categoria il servizio *doganale*, quello dei *porti* e *fari* e non pochi rami attinenti alla *difesa della linea di confine*, salva per tutti la parte che si riferisce all'Amministrazione centrale, per cui figureranno invece nel terzo gruppo. Nella seconda categoria porremo i servizi, disseminati nel territorio, dell'Amministrazione giudiziaria, della polizia interna, dell'istruzione pubblica, delle poste e telegrafi, ecc., che hanno un rapporto più stretto colla superficie o colla popolazione, ammesso che la quantità di affari, di reati, di corrispondenze, ecc., sia in media eguale in entrambi gli Stati per ogni chilometro quadrato di superficie o per ogni 1000 abitanti. Infine apparterrà alla terza categoria, come quello che presumibilmente deve crescere in una ragione intermedia delle precedenti, il personale delle Amministrazioni centrali, in quanto appunto i rapporti, che fanno capo al centro, derivano o da servizi pubblici sviluppantisi in ragion di periferia o da servizi sviluppantisi in ragion di superficie. Oltre a ciò vuolsi riflettere, che se gli affari devoluti ai dicasteri centrali sono cinque o sei volte più numerosi in un paese che in un altro, non ne viene che il primo debba irreggimentare un numero quintuplo o sestuplo di funzionari, ma uno forse quadruplo soltanto, compensandosi la differenza colle minori intermittenze del lavoro, colla ripartizione più razionale di questo, coll'unità di indirizzo e di rappresentanza, ecc.

Seguendo tali criteri per la formazione di gruppi scelti, siamo giunti dopo lunga elaborazione degli elementi di confronto, a questi risultati approssimativi (2).

---

(1) Se le aree di due Stati circolari stanno fra loro nel rapporto di 1 a 7, le periferie o linee di confine stanno invece fra loro come 1 a 2,65 circa. Le corrispondenti proporzioni dovrebbero dunque trovarsi nella quantità di personale adibito ai servizi diffusi e ai servizi periferici.

(2) Gli elementi di confronto sono desunti da una pubblicazione della Direzione generale della Statistica sui *Ruoli organici delle Amministrazioni civili e militari del Regno al* 1° *luglio* 1891 *confrontati con quelli degli antichi Stati italiani al* 1° *gennaio* 1859 (*Annali di Statistica*, serie 4ª, n. 62) e dal *Compendio degli organici delle Amministrazioni civili e militari dello Stato al* 30 *giugno* 1888 (*Annali di Statistica*, serie 4ª, n. 28), che contiene pure i dati del 1883. Per l'elaborazione e comparazione veggasi il nostro scritto: *La Burocrazia di Stato in Italia dal* 1859 *al* 1891 (*Riforma sociale*, 1895).

SERVIZI CIVILI.

ANNI	Periferici o di confine (Servizio doganale)		Centrali		Diffusi o non altrimenti classificati	
	Personale (num.)	Spesa per stipendi (in lire)	Personale (num.)	Spesa per stipendi (in lire)	Personale (num.)	Spesa per stipendi (in lire)
1859	21.080	11.922.204	5.068	9.877.562	49.922	56.404.550
1883	18.134	17.645.700	3.170	9.245.000	55.330	99.941.565
1891	18.215	18.986.090	3.409	10.009.714	65.992	122.544.034

Dal prospetto risulta chiaro che se in complesso vi fu aumento nel numero dei funzionari dal 1859 al 1883 e, più ancora, al 1891, tale aumento è imputabile ai servizi diffusi e si spiega col crescere della popolazione e coll'estendersi della attività governativa; laddove nei servizi periferici e centrali si è realizzata quella economia di personale, che era da prevedersi anche per considerazioni *a priori*. Circa la spesa complessiva, l'aumento dal 1859 al 1883 appare dipendente quasi per intero dalla elevazione dello stipendo medio; dal 1883 al 1891 deriva invece in massima parte dall'aumento numerico del personale.

§ **4.** Le differenze di legislazione dei diversi Stati e conseguentemente i diversi criteri di classificazione adottati dagli uffici di Statistica rendono, come è noto, assai difficili i confronti internazionali. Tutto quel che di generale si può dire in proposito è che, tenute ben presenti le differenze di nomenclatura e di comprensione delle singole *voci*, si formino anche qui gruppi scelti degli elementi che più si somigliano nelle statistiche confrontate e i restanti si raccolgano in classi a larghi confini, sì che molte discordanze parziali vi si possano considerare compensate. Naturalmente molto dipende dall'avvedutezza e spirito critico di chi fa la comparazione.

L'Istituto internazionale di Statistica a questo proposito ha spiegato una attività degna del maggior encomio, non solo nel raccomandare alle pubbliche Amministrazioni dei diversi Stati formularii e metodi di rilievo intesi ad agevolare i confronti, ma altresì nel coordinare i materiali esistenti in forma di saggi di statistiche comparative. Il *Bollettino*, che esso pubblica regolarmente, contiene memorie, relazioni e comunicazioni sempre interessanti, qualche volta costituenti dei veri modelli di metodologia statistica (1).

(1) Veggasi nel *Bulletin de l'Institut international de Statistique*, Tome XV 1^{ère} Livraison, l'elenco generale delle memorie, relazioni, comunicazioni e voti pubblicati dall'epoca della fondazione dell'Istituto (1885) fino alla Sessione di Londra (1905).

Una delle questioni più importanti trattate nelle ultime sessioni dell' Istituto fu quella del metodo maggiormente raccomandabile, dal punto di vista statistico-internazionale, per la compilazione dei *Bilanci delle Società anonime*. Su di essa riferì con molta competenza Alfredo Neymarck, che era al tempo stesso relatore della Commissione extra-parlamentare nominata in Francia per lo studio delle riforme da introdurre nella legislazione delle società per azioni. Il questionario preparato per l' Istituto concerneva gli elementi che debbono figurare nello attivo e nel passivo, le condizioni alle quali l'eccedenza dell'attivo sul passivo costituisce un beneficio ripartibile tra i soci alla fine dell'esercizio; il modo di valutazione dei materiali, delle merci, degli immobili e dei titoli; il modo di ammortizzo delle spese di costituzione della società, delle spese di primo impianto; la specificazione delle riserve create nel corso degli esercizi sotto i nomi di *riserva speciale*, *riserva straordinaria*, *fondo di previdenza*, ecc.; la possibilità di stabilire un tipo uniforme di bilancio per tutte le società, ecc. Il questionario provocò due Memorie assai interessanti, l'una di E. Léautey, direttore dello *Institut comptable* di Parigi, l'altra di J. Claudel, professore all'*Association Polytechnique*. La brevità non consente di qui riprodurre le loro considerazioni in materia così delicata e variamente regolata dalle leggi dei diversi paesi, nè i modelli di bilancio proposti; il lettore può vederli nei Rendiconti della IX sessione dell' Istituto (1).

(1) Veggasi *Bulletin*, ecc., Tome XIV, 2ème Livraison. — La Memoria del Léautey fu anche pubblicata a parte col titolo: *L'unification des Bilans des Sociétés par actions*, Paris Librairie comptable et administrative.

APPENDICE AL LIBRO I.

DEL CENSIMENTO

(In particolare dell'ultimo censimento italiano).

Sommario. § 1. Importanza scientifica e politico-amministrativa del censimento. — § 2. Caratteri principali del censimento. — § 3. Notizie che ne formano oggetto. — § 4. Modo di raccolta delle notizie. — § 5. Organi esecutivi del censimento. — § 6. Scopo immediato di questa operazione statistica. — § 7. Osservazioni critiche sull'ultimo censimento italiano.

§ **1.** La grande importanza che ha il *Censimento della popolazione* ci induce a dare qualche cenno di questa operazione statistica in appendice alla parte che tratta della Statistica come forma di osservazione.

Per « stato della popolazione » intendiamo la consistenza numerica di questa, considerata in un dato momento, in un dato territorio, nelle principali sue forme di aggruppamento o di coesione e nei principali caratteri sociali, per cui si distinguono gli individui che la compongono.

Per « movimento » intendiamo i fenomeni che inducono variazioni nella consistenza numerica della popolazione, nelle sue forme d'aggruppamento o di coesione e nelle proporzioni delle singole classi distinte per caratteristiche demografiche determinate.

Il movimento della popolazione è rilevato soprattutto col mezzo dei ***registri dello stato civile*** (nascite, matrimoni, legittimazioni, morti) e dei *registri dei passaporti* (emigrazione); lo stato, principalmente col mezzo del *censimento.*

L'importanza del censimento è grande così dal punto di vista scientifico, come da quello politico-amministrativo. Scientificamente, perchè moltissimi dati statistici non acquistano valore o significato definito, se non riferiti ad un termine naturale di paragone, quale appunto la popolazione in seno alla quale si producono i fenomeni osservati. Nei riguardi politico-amministrativi, perchè il censimento fornisce criteri per determinare l'ordine gerarchico di certi istituti, la composizione delle assemblee deliberanti, l'estensione di molti servizi pubblici, ecc. Non meno di 27 leggi in Italia hanno come base d'applicazione il censimento (1); tali, la *legge elettorale politica*, che ripartisce il numero dei

(1) Veggasi: *Studi e proposte per l'esecuzione del IV censimento generale della popolazione del Regno* (Direz. gen. della Statist., Roma 1900, pag. 49-53).

deputati secondo la popolazione delle provincie e dei collegi; la *legge comunale e provinciale*, che stabilisce un numero vario di consiglieri e di assessori a seconda del numero degli abitanti dei comuni e analogamente dispone per la composizione dei Consigli e delle Deputazioni provinciali; la *legge sulla pubblica sicurezza*, che istituisce uffici di Questura nelle città con più di 60.000 abitanti; quella sull'*Ordinamento della pubblica istruzione*, che classifica i ginnasi, i licei, le scuole tecniche e gli istituti tecnici; quella sul *dazio di consumo* che distingue i comuni in aperti e chiusi e i chiusi in parecchie classi con diverse tariffe: e così via. Rientra nel novero anche la legge 2 gennaio 1898 che approvò la *Convenzione monetaria* 29 ottobre 1897, statuente l'impegno degli Stati della Lega latina (Italia, Francia, Svizzera, Belgio e Grecia) di non coniare più di 7 lire in monete divisionarie d'argento, per abitante.

Si parte insomma sempre dalla presunzione che la somma degli interessi da regolare e da invigilare o il numero degli individui capaci siano in un certo rapporto coll'importanza dei gruppi e centri di popolazione.

In mancanza di un censimento si può calcolare la popolazione di un paese coi dati del movimento dello stato civile (come si disse a proposito dei metodi di Halley e di Hermann per le tavole di sopravvivenza) e con quelli dell'emigrazione, dell'immigrazione o almeno dei rimpatrii; ma soprattutto per l'imperfetta conoscenza dei movimenti migratorii, gli errori sarebbero inevitabili e talvolta assai considerevoli. I registri comunali di anagrafe potrebbero, se tenuti in buon ordine, supplire alla mancanza di un vero e proprio censimento; nel fatto però, almeno in Italia, la loro utilità non va oltre l'ordinario bisogno della compilazione dei ruoli di contribuenti, delle liste di elettori e di giurati, ecc. Mentre il censimento colla sua simultaneità d'esecuzione, colla regolare sua periodicità, col rilievo esteso a tutte le classi di popolazione ed a caratteri, la cui notizia non si saprebbe attingere ad altre fonti, esclude l'arbitrio inevitabile delle congetture, che sono a base dei calcoli, e serve a rimettere a giorno gli stessi registri anagrafici.

§ **2.** I caratteri principali del censimento sono: la *periodicità*, la *simultaneità* e la *universalità*.

In Francia, Germania e Svezia il censimento si ripete ogni cinque anni; in molti altri paesi ogni dieci, in altri ancora a periodi non prestabiliti. Da noi la popolazione fu censita nel 1861, 1871, 1881 e 1901, saltandosi, per ragioni di economia, il 1891, sebbene le leggi che ordinarono i censimenti del 1871 e 1881 accennassero a questa, come ad operazione ripetibile ogni decennio.

La simultaneità implica che i rilievi si compiano a giorno fisso per tutto il territorio dello Stato. Si può discutere sulla scelta del giorno. I tre primi censimenti italiani ebbero luogo al 31 dicembre; per il quarto fu preferita la data del 9-10 febbraio. Motivi del cambiamento

furono: che alla fine dell'anno, per la ricorrenza delle feste natalizie e di capo d'anno, molte persone, che vivono abitualmente lontane dai loro parenti, fanno ritorno in famiglia; in quell'epoca si accentua la emigrazione temporanea da certe provincie per certe altre; così che il censimento non ritrarrebbe la popolazione nelle sue sedi abituali. I giorni festivi sono anche meno adatti pel disbrigo del grande lavoro iniziale di raccolta delle schede e di corrispondenza tra l'ufficio centrale e gli enti locali.

In Inghilterra la data preferita è la prima domenica d'aprile; in Francia i censimenti del 1891, 1896 e 1901 ebbero luogo rispettivamente, il 12 aprile, il 29 marzo e il 14 aprile; Germania e Svizzera hanno adottato il 1° dicembre; gli Stati del Nord America, il 1° giugno, ecc.

L'Istituto internazionale di Statistica, in una delle sue ultime riunioni aveva formulato il voto, confermato anche dal Congresso di Pietroburgo, che i diversi paesi civili eseguissero il censimento in modo simultaneo nel 1900. Il voto non fu accolto dai Governi. Questa simultaneità internazionale avrebbe per altro oggi dei vantaggi più apparenti che reali; quel che davvero importerebbe, sarebbe piuttosto la uniformità di criteri nel condurre queste colossali operazioni. Importerebbe pel riscontro dei dati da paese a paese uno stesso metodo di rilevazione dei presenti occasionali e degli assenti temporanei; che la nazionalità attribuita da alcuni Stati, specie del Sud America, agli immigrati dopo un certo tempo di dimora, non fosse in contraddizione colla nazionalità che essi molte volte conservano, secondo la legge del paese d'origine, e così via.

La *universalità*, come caratteristica dei censimenti moderni, vuolsi intendere nel senso che nessuna classe di persone e nessuna parte del territorio siano escluse dai rilievi. Questioni a tale riguardo oggi più non si fanno; si può discutere dell'opportunità di fare un censimento più specializzato per le città che per le campagne; ma niente più. In altri tempi invece si limitarono i censimenti agli uomini liberi, escludendosi gli schiavi, o agli individui adulti, escludendosi i bambini.

§ **3.** Le notizie che sogliono formare oggetto del censimento riguardano i *luoghi di dimora*, la ***qualità delle convivenze*** e i ***singoli individui***.

A) La identificazione topografica, per dir così, dei censiti deve anzitutto risultare dal nome del comune e mandamento, da quello della frazione e sezione ed eventualmente anche della parrocchia; infine dal nome della via, piazza o casale e dal numero assegnato alla casa, la quale può appartenere ad un centro di case contigue o rientra nel novero delle case sparse. Questa distinzione, affatto relativa, ha per risultato di classificare la popolazione in agglomerata e sparsa. Per *centro* s'intende una aggregazione di case separate da strade, ove sogliono convenire gli abitanti dei luoghi vicini per ragioni d'affari, di culto e simili. Un gruppo di poche case in una valle appartata, ove convengono

gli abitanti delle case sparse intorno può, per le ragioni dette, essere considerato come un centro di popolazione; laddove lo stesso gruppo di case non si dirà più un centro, in vicinanza di borgate o città.

Dall' unità « casa » si suol discendere con ulteriore specificazione all'*abitazione*, individuata secondo il piano al quale si trova, secondo il numero e la destinazione delle stanze, ecc.

B) Le forme di coesione, cui può dar luogo una popolazione, sono numerosissime. La Demografia le raccoglie in tre vasti gruppi: forme di coesione per simpatia, per coordinazione gerarchica e per divisione del lavoro. Ogni individuo può appartenere ad un tempo a svariati nuclei sociali: la famiglia, la comunità religiosa, la scuola, il partito politico, l'impresa industriale, l'associazione letteraria, artistica o scientifica, ecc., ecc. Il censimento non tien conto che delle forme di coesione o di aggruppamento che implicano una vera e propria convivenza temporanea o duratura. Quindi le *famiglie* propriamente dette (focolari domestici); *alberghi* e *locande*, *caserme*, *collegi*, *convitti* ed *educandati*, *ospizi*, *ospedali*, *brefotrofi*, *carceri*, ecc.

C) Le notizie individuali concernono soprattutto lo *stato civile* (in largo senso) e lo *stato professionale* dei censiti; in via accessoria, la loro capacità a leggere e scrivere, i loro difetti apparenti (cecità, sordomutismo, talvolta l'idiozia e la pazzia) ed alcune altre caratteristiche.

Lo stato civile in largo senso degli individui risulta dalle indicazioni concernenti il cognome, il nome e la paternità, il sesso, l'età, la condizione di celibe (o nubile), coniugato, vedovo, divorziato, il luogo di nascita, la cittadinanza, la residenza (dimora abituale) e la dimora occasionale, l'appartenenza ad una confessione religiosa. Anche questa ultima indicazione, non riferendosi alla fede in cui si crede, ma all'ambiente in cui si è nati, serve a definire lo stato civile degli individui.

Lo stato professionale si determina colle notizie riguardanti la professione unica o principale del censito, la professione od occupazione accessoria, il posto gerarchico nel ramo di produzione o di commercio, l'ambiente di esercizio del lavoro, e la condizione eventuale di disoccupazione. Come complemento di queste indicazioni si ha spesse volte la domanda relativa alla proprietà di terreni o fabbricati.

§ 4. La raccolta delle notizie si fa per mezzo di *schede di famiglia* o per mezzo di *bollettini individuali* o con *sistema misto*. Il nostro censimento del 1881 fu eseguito col metodo delle schede di famiglia, contenenti le dichiarazioni dei capi-famiglia, scritte direttamente da essi o raccolte dalla loro viva voce, quando erano illetterati, a cura dei commessi del censimento. Gli uffici comunali trascrissero poi le notizie concernenti ciascun individuo, riportandole dalle schede di famiglia su appositi bollettini individuali, il cui spoglio ebbe luogo presso la Direzione generale della Statistica in Roma.

Questo sistema fu preferito allora per due ragioni: 1° per dar modo ai comuni di correggere e sistemare i loro registri d'anagrafe, mediante le schede di famiglia, che rimanevano a loro disposizione; 2° perchè si riteneva disadatto per una popolazione semianalfabeta il metodo del bollettino individuale, il quale richiede che si riempiano tanti foglietti distinti, quanti sono i membri della famiglia. Il fatto ha smentito il primo argomento; ben pochi furono i comuni, che del censimento del 1881 profittarono per riordinare i loro registri d'anagrafe. Inoltre, la trascrizione delle notizie dalle schede di famiglia su bollettini individuali non avvenne senza grave spesa e perdita di tempo, nè senza errori di copiatura, che resero necessario un considerevole lavoro di corrispondenza tra la Direzione generale della Statistica e i Comuni.

La seconda ragione addotta nemmeno reggeva; la scheda di famiglia essendo per necessità un foglio di grande formato, con molte colonne orizzontali e verticali fra cui è facile lo scambio, riesce ben più disadatta alle popolazioni, tra le quali poco è diffusa la coltura elementare, che non riesca il bollettino individuale, assai più semplice, comodo e chiaro. Opportunamente quindi il censimento del 9-10 febbraio 1901 si attenne al sistema dei *bollettini individuali*, secondo l'esempio della maggior parte degli Stati, e della *busta di riepilogo*, in cui essi vengono racchiusi, la qual busta funge in certo modo da scheda di famiglia, semplificata. Infatti a tergo della medesima il capo-famiglia doveva dare l'elenco delle persone inscritte nei bollettini, indicando le presenti con dimora abituale od occasionale e le assenti temporaneamente dalla famiglia.

§ 5. Veniamo agli *organi esecutivi del censimento*. Da noi il censimento ultimo si eseguì in ogni comune sotto la direzione e responsabilità del *sindaco*, assistito da una *Commissione di censimento* formata dalla Giunta municipale e da altre persone in numero eguale a quello degli assessori effettivi. Tali persone vennero scelte di preferenza tra i funzionari in attività di servizio o a riposo, tra i parroci, i medici, i maestri, ecc. La Commissione coadiuva il sindaco nella revisione preliminare dei nomi delle vie, numerazione dei fabbricati, divisione del territorio comunale in frazioni e sezioni di censimento; nella scelta dei commessi; nella revisione e correzione delle schede, ecc.

Il verificare se i comuni hanno provveduto in tempo per la nomina delle Commissioni e per la scelta dei commessi, se furono osservate le norme del regolamento per la divisione del territorio in frazioni, ecc., e il rivedere i prospetti eseguiti dagli uffici comunali in base allo spoglio delle notizie relative alle case, abitazioni e famiglie e alla popolazione presente e residente, agglomerata e sparsa, spetta alle *Giunte provinciali di Statistica*, composte di otto membri nominati per metà dal Consiglio provinciale, per metà dal prefetto, che li sceglie tra i professori

d'Economia e di Statistica, gli ingegneri del Genio civile, i provveditori ed ispettori scolastici.

L'alta direzione e sorveglianza su tutte le operazioni del censimento spetta al Ministero di agricoltura, industria e commercio, e per esso alla Direzione generale della Statistica. A questa vengono trasmesse le schede individuali nelle rispettive buste di riepilogo; ed essa è incaricata di tutte le altre operazioni di spoglio e della pubblicazione dei risultati.

Le operazioni preliminari del censimento comprendono pure l'invio in tempo opportuno da parte dei comuni alla Direzione generale della Statistica di prospetti indicanti il numero probabile delle famiglie e degli abitanti nel comune, secondo i registri d'anagrafe, e ciò allo scopo di agevolare la distribuzione degli stampati (schede individuali e buste di riepilogo) in quantità sufficiente. Inoltre, dieci giorni prima di quello fissato pel censimento, i commessi si recano in ciascuna delle abitazioni della sezione loro assegnata, per prender nota, fra le altre cose, del numero approssimativo dei membri della famiglia, a cui saranno rimesse poi altrettante schede individuali. Sicchè il censimento vero e proprio può dirsi preceduto da una duplice rilevazione approssimativa.

Da noi, la distribuzione delle schede accompagnata da istruzioni scritte e verbali, fu eseguita nei giorni dal 6 al 9 febbraio 1901: il ritiro, previa verifica ed eventuali correzioni da parte degli stessi commessi, avvenne nei giorni dal 10 al 12.

§ **6.** Scopo immediato del censimento è di accertare la ***popolazione di fatto*** o ***presente*** (ossia il numero delle persone presenti per qualsiasi motivo in ogni comune alla data del censimento) e la popolazione ***residente*** o ***legale*** o ***di diritto*** (come anche dicesi) la quale risulta dall'addizione dei ***presenti con dimora abituale*** nel comune in cui sono censiti, cogli ***assenti temporaneamente*** dal comune stesso.

Per l'applicazione delle leggi, che fanno riferimento al numero degli abitanti, si prende per base la popolazione ***residente*** e non la ***presente***, la prima rappresentando meglio le condizioni demografiche normali dei diversi comuni. Se infatti si prendesse per base la popolazione presente, potrebbe avvenire che un piccolo comune noverasse alla data del censimento un numero di abitanti molto superiore all'ordinario (per esempio, a causa di una dislocazione di truppe o di operai immigrati per la costruzione di una ferrovia, ecc.) e vantasse percio il diritto ad avere una rappresentanza municipale più numerosa o a divenir sede di qualche ambito ufficio governativo, il che non sarebbe giusto.

Secondo il sistema da noi adottato, l'assente temporaneamente dalla famiglia o dal comune, ma presente nel Regno, doveva dare origine non ad una sola, bensì a ***due*** schede facentisi contropartita: l'una, scritta da lui stesso nel luogo di sua dimora occasionale; l'altra, scritta

dal suo capo-famiglia, che lo indicava assente temporaneamente dal comune di abituale dimora. Sicchè l'assente temporaneo era tenuto a rispondere in modo affermativo alla seconda parte del 3° quesito (veggasi innanzi, il modello di scheda del censimento 9-10 febbraio 1901), sottolineando la parola « occasionale », laddove il suo capo-famiglia doveva in altra scheda rispondere per lui affermativamente al quarto quesito.

Naturalmente, disposizioni speciali furono fissate per casi particolari. Così le persone che alla data del censimento si trovavano in ferrovia, dovevansi censire nel luogo d'arrivo nel giorno 10 febbraio 1901; gli impiegati postali in servizio notturno eran contati come presenti nella famiglia in cui rientravano dopo compiuto il turno di servizio, ecc.

I bambini dati a balia si considerarono presenti con dimora occasionale presso la famiglia, cui erano affidati; mentre la lor famiglia naturale doveva indicarli come assenti temporanei. Essi perciò davano origine a due schede ciascuno. Lo stesso ripetasi per gli studenti posti in un convitto o collocati per ragion di studi presso qualche famiglia; degli infermi (non cronici) negli ospedali, degli ospiti in visita, dei viaggiatori negli alberghi e locande, dei detenuti sotto processo e dei condannati per un tempo minore di sei mesi. Invece i fanciulli a carico dell'assistenza pubblica si consideravano come aventi dimora abituale presso la famiglia, ove si trovavan collocati a baliatico; i ricoverati in asili di cronici, manicomii, ospizi di mendicità, ecc., come aventi dimora abituale in detti stabilimenti; i condannati a più di un anno come aventi dimora abituale nella casa di pena; le guardie accasermate, i militari in servizio effettivo, i domestici coabitanti col padrone, come aventi dimora abituale nel luogo dove si trova la caserma o il quartiere o la famiglia del padrone. Per queste persone la scheda era dunque unica e non duplice; le loro famiglie naturali non dovevan considerarle come assenti temporanei.

§ **7.** *Osservazioni critiche sulla scheda dell'ultimo censimento italiano.* — Le 19 domande rivolte ai censiti erano così formulate nella scheda:

1. Cognome, nome, paternità.....
2. Relazione di parentela o convivenza col capo di famiglia.....
3. Presente con dimora **abituale** — **occasionale.**
4. **Assente temporaneamente** dalla famiglia — Luogo dove trovasi l'assente.
5. Sesso: **maschio** — **femmina.**
6. Anno di nascita..... mese.....
7. Luogo di nascita — Chi è nato nel Regno dica in quale comune..... e in quale provincia..... — Chi è nato all'estero dica in quale Stato.....

8. Chi non è cittadino italiano, dica a quale Stato appartiene.....
9. **Celibe - Nubile — Coniugat... Vedov...**
10. **Sa leggere — non sa leggere.**
11. Religione. Chi appartiene ad un culto, dica qual è.....
12. Chi ha intestati al suo nome in catasto o nei ruoli delle imposte beni immobili, dica se ha **terreni** — se ha **fabbricati.**
13. Condizione o professione unica o principale.....
14. Chi esercita l'agricoltura, dica se **conduce** o **lavora terreni proprii** (o della famiglia) ovvero se è **fattore** — **fittaiuolo** — **enfiteuta (utilista)** — **colono** o **mezzadro** — **contadino obbligato** — **giornaliero (bracciante di campagna).**
15. Chi è occupato in un'industria o in un commercio, dica se è **padrone** — **direttore** — **capotecnico** — **impiegato** — **commesso** — **agente** — **viaggiatore** — **artigiano indipendente** — **operaio** — **facchino** — **bracciante.**
16. L'operaio e in generale chi esercita un lavoro manuale, dica se lavora in un **opificio** o altro **locale** del padrone, ovvero **nella propria abitazione.**
17. L'operaio, artigiano, domestico o bracciante, che sia attualmente **disoccupato,** dica da quanto tempo..... e se **per malattia** ..... o **per altro motivo.**
18. Professione od occupazione accessoria.....
19. È **cieco** — è **sordomuto.**

Da una relazione del dottor J. Bertillon all'Istituto internazionale di Statistica (Sessione di Budapest) togliamo alcune notizie comparative sugli ultimi censimenti dei paesi d'Europa.

Il 1°, 2°, 5°, 6°, 7° e 9° sono quesiti proposti nei censimenti di tutti i paesi, salve alcune varianti o aggiunte. Così in parecchi paesi si domanda l'*età*, in altri l'*anno di nascita*, in altri ancora la *data di nascita* (il primo modo, come sappiamo, lascia maggiormente aperto l'adito alle denunzie di età in cifre rotonde). Dove vige il divorzio si suol chiedere, a proposito dello stato civile, la condizione di *divorziato*; in Olanda e nel Portogallo si distinguono anche i *separati legalmente*.

I quesiti 3° e 4° sono invece formulati nei modi più varii. Nei censimenti britannici anzi non son nemmeno posti, il che vuol dire che si tien conto della sola popolazione di fatto o presente; in Francia e Portogallo il censito deve dire se egli ha o non ha nel comune la residenza abituale; in Isvizzera, da quanto tempo egli è assente dalla famiglia, e in Ungheria, qual'è la causa dell'assenza.

La *nazionalità* o *cittadinanza* non è domandata in Irlanda, Danimarca e Russia; lo è per le sole persone nate fuori dei territori britannici in Inghilterra e Scozia.

Il quesito della *conoscenza del leggere e dello scrivere* figura nei censimenti dell'Irlanda, Russia, Austria, Belgio, Francia, Portogallo, Bulgaria e Serbia; figurava, ma non figura più in quelli della Danimarca,

Norvegia, Svizzera e Olanda, dove la coltura elementare è così diffusa da rendere omai superflua la domanda. Questa è taciuta pure in Inghilterra e Scozia. Da noi ora è limitata alla condizione del ***saper leggere***, mentre nel censimento del 1881 si estendeva al ***saper scrivere;*** altrove si propongono distintamente le due questioni.

Inghilterra, Scozia, Belgio, Francia e Portogallo non pongono il quesito della religione.

Francia, Danimarca, Olanda e Portogallo non rilevano la *lingua parlata.*

Da noi il quesito non figura sulla scheda; tuttavia fu stabilito che i comuni, nei quali si trovano gruppi di famiglie di nazionalità italiana parlanti abitualmente un dialetto francese o tedesco o slavo o greco o albanese o catalano, compilassero in base alle notizie raccolte dai commessi un prospetto numerico di tali famiglie.

Intorno alle ***infermità apparenti*** tacciono i censimenti della Svizzera, Belgio e Portogallo, cui si aggiungono ora l'Austria e l'Olanda. Degli altri paesi, quali si limitano alla cecità e sordomutismo, quali si informano anche dell'idiozia e della pazzia.

Qua e là troviamo proposti alcuni altri quesiti d'importanza demografica, come la *durata del matrimonio* o *della vedovanza*, il ***numero dei figli viventi***, ecc.

Difficile è riassumere le differenze tra i quesiti che concernono lo ***stato professionale*** dei censiti nei diversi paesi. Tutti domandano la professione principale e la condizione gerarchica; parecchi, tra cui il nostro, la professione accessoria e lo stato di disoccupazione; ma sempre con varianti degne di nota. Così si chiede in Danimarca, Austria, Ungheria, Svizzera e Francia la ditta o ragion sociale e la specie dell'industria; in Francia ed in Ungheria, anche il numero degli operai. Il piano di classificazione delle professioni è diversamente specializzato. In Francia, da sole 60 voci (censimenti del 1886 e 1891) si passò addirittura a 1320. In Inghilterra l'elenco ne comprende 400; in Germania ed Austria, 200; da noi, 300 circa.

In Germania si eseguisce il censimento delle professioni e degli esercizi industriali in modo separato dal censimento demografico in senso stretto; il primo si ripete a intervalli di dieci anni e più, il secondo a quinquennii, la brevità del periodo essendo richiesta dalla necessità di rivedere la ripartizione dei proventi doganali e delle quote di concorso alle spese militari fra i diversi Stati dell'Impero in ragion del numero degli abitanti.

I criteri di classificazione delle professioni non collimano con quelli che sono raccomandabili per una statistica delle industrie. Là si ha riguardo al *genere di lavoro*, ossia al modo col quale si applicano alla produzione i singoli individui; qui si ha di mira piuttosto il ***risultato*** finale, ossia la specie di prodotto ottenuto. Per una statistica industriale importa riunire per ciascun opificio i dipendenti di un medesimo

imprenditore, sebbene specificati secondo i generi di lavoro, il grado gerarchico od altro; quindi, ad esempio, in una filatura di cotone gli operai filatori, i macchinisti addetti ai motori, i fuochisti, i facchini, ecc., figurano come unità di un medesimo gruppo; invece per una statistica professionale i macchinisti, siano essi addetti a stabilimenti di filatura o di tessitura, o ad officine meccaniche, formeranno una categoria a parte; e così i fuochisti, i facchini, ecc.

Le principali categorie e classi fissate nel nostro ultimo censimento sono le seguenti:

A) *Agricoltura.*

(Agricoltura propriamente detta, silvicoltura e allevamento del bestiame, pesca e caccia).

B) *Industria.*

(Industrie estrattive; industrie mineralurgiche, metallurgiche e meccaniche; lavorazione delle pietre, argille e sabbie; industria edilizia; fabbricazione di prodotti chimici; lavorazione del legno e della paglia, e fabbricazione di mobili di legno e di utensili domestici; industrie della carta, tipografiche e litografiche; industrie tessili; lavorazione delle pelli e di altri prodotti animali; industrie attinenti al vestiario e all'acconciatura della persona; costruzione di veicoli; industrie di precisione e di lusso; industrie alimentari; operai non specificati).

C) *Commercio.*

(Trasporti per acqua, su strade ordinarie e ferrovie; poste, telegrafi, telefoni; vendita di merci e derrate all'ingrosso e al minuto; credito e cambio, assicurazione, mediazione pubblica e privata, commissioni, rappresentanze commerciali; esercizi pubblici).

D) *Persone addette ai servizi domestici e di piazza.*

(Persone addette al servizio domestico; persone addette ai servizi di piazza e affini).

E) *Professioni ed arti liberali.*

(Amministrazione pubblica; amministrazioni private; difesa del paese; insegnamento; culto; professioni sanitarie; professioni legali; lettere e scienze applicate; arti belle).

F) *Persone non occupate in alcuna professione.*

(Persone che vivono specialmente di reddito; persone mantenute dalla famiglia; persone assistite dalla carità pubblica e privata o viventi a carico dello Stato).

G) *Persone di professione o condizione ignota.*

Quanto ai risultati del nostro ultimo censimento, comparati con quelli degli anteriori nostri e con alcuni stranieri, ci conviene rinviare

il lettore alla pubblicazione citata in nota (1). Qui ci limiteremo a poche osservazioni di metodo.

A nostro avviso sarebbe stata desiderabile una maggior semplificazione della scheda, concentrando l'inchiesta su due obbietti: lo *stato civile* (in largo senso) e lo *stato professionale* d'ogni individuo *presente* e intensificando per converso il lavoro di spoglio col combinare fra loro i quesiti due a due, tre a tre, se non in tutti i modi possibili (ciò che sarebbe soverchio), almeno in tutti i modi più interessanti. Criterii giustificativi i seguenti: Quanto meno colto è un paese, tanto più semplici e chiare debbono essere le domande che gli si rivolgono colla scheda; quanto più esso è gravato d'imposte e sospettoso del fisco, tanto meno opportuno riesce il chiedergli cose allusive allo stato economico delle persone. Dove, come da noi, la popolazione è molto migrante, dalla campagna alla città, dal monte al piano, dall'interno all'estero, ivi importa togliere ogni possibile equivoco fra dimora abituale e dimora occasionale e non complicare queste domande con quella dell'assenza temporanea o non temporanea. Alcune caratteristiche rare, sporadiche, conoscibili per altra via, non dovrebbero formare oggetto di una speciale interrogazione rivolta all'universalità dei cittadini con danno dell'economia del questionario. Infine l'utilità scientifica di certi quesiti si riduce a ben poca cosa, quando essi vengono proposti solo sulla scheda del censimento, mentre non si riproporranno nei rilievi annuali del movimento della popolazione.

Siffatti criterii avrebbero potuto decidere al sacrificio dei quesiti concernenti la religione, la proprietà immobiliare, la cecità e il sordomutismo. Invero, qualora nelle statistiche annuali dei matrimoni, delle nascite, delle legittimazioni, delle morti, dell'emigrazione, dei riformati alle leve, ecc., si registrassero i casi secondo l'appartenenza di culto, secondo la condizione di proprietario o non-proprietario, ecc., apparirebbe la ragion decisiva dello stabilire analoghe partite nel censimento: basterebbe infatti mettere in rapporto il movimento annuale di quei fatti demografici collo stato della popolazione, per categorie omonime, per ottenere preziose informazioni sulle differenze di nuzialità, di fecondità, di migrabilità, di difetti esimenti dal servizio militare, ecc., che caratterizzano gli israeliti in confronto degli evangelici e dei cattolici, i proprietari di stabili in confronto dei non-proprietari, e così via. Ma dal momento che siffatte domande non si ripresenteranno nella rilevazione dei fenomeni di movimento, non possiamo sperarne che semplici dati descrittivi e per molte ragioni malsicuri, mentre per altra via se ne avrebbero di più attendibili.

---

(1) Veggasi *Censimento della popolazione del Regno d'Italia al 10 febbraio 1901, vol. V, Relazione sul metodo di esecuzione e sui risultati del censimento raffrontati con quelli dei censimenti italiani precedenti e di censimenti esteri* (Direzione generale della Statistica, Roma 1904).

Infatti, circa il quesito della religione (formulato in maniera dubbia, che si prestava ad interpretarlo o come religione in cui *si crede* o come religione in cui *si è nati*) avremmo potuto contentarci di una indagine speciale — fuori censimento — fatta col concorso dei rabbini e dei pastori evangelici, per determinare il numero dei seguaci delle rispettive chiese. Così, pur senza spogliare da capo a fondo i ruoli delle imposte dirette, eliminando le intestazioni ripetute allo stesso nome, il numero approssimativo dei proprietari d'immobili poteva ottenersi dalle statistiche annuali delle *successioni*, previi alcuni necessari adattamenti. Infatti uno spoglio dei ruoli, colla debita unificazione delle intestazioni ad un medesimo nome, si fa ogni anno per i proprietari *defunti;* se l'Amministrazione finanziaria ci facesse conoscere di costoro la classificazione per età, sesso e stato civile, potremmo calcolare il numero dei proprietari *viventi*, meglio e con dettagli più interessanti che non possiamo aspettarci dal censimento (1).

Alla vagheggiata semplificazione della scheda avrebbe poi grandemente contribuito un diverso modo di accertare la popolazione legale o residente. Le domande: *presente con dimora abituale — con dimora occasionale — assente temporaneamente dalla famiglia — luogo dove trovasi l'assente* — sono complicate e dall'esperienza dimostrate causa di facili equivoci. Come già dicemmo, ogni presente con dimora occasionale in un comune del Regno doveva dare origine, non ad una, ma a due schede, la prima scritta *da lui* nel luogo di dimora occasionale, la seconda scritta *per lui* dal suo capo-famiglia, che lo indicava come

---

(1) Nel 1881 il censimento diede 2.733.000 maschi proprietari di terreni o fabbricati. Lo spoglio dei ruoli per agenzie nel 1879 in occasione degli studi per la riforma della legge elettorale politica, spoglio in cui vennero eliminate le intestazioni ripetute allo stesso nome nel distretto proprio di ogni agenzia, diede per risultato ben 5 milioni di contribuenti maschi maggiorenni inscritti nominativamente nei ruoli dell'imposta sui terreni e di quella sui fabbricati, senza contare i contribuenti compresi in articoli intestati a più comproprietari. Ora, data l'estensione delle agenzie (che comprendono in media 13 comuni, ossia 450 chilometri quadrati di superficie), il numero dei proprietari di immobili situati in distretti di due o più diverse agenzie, non deve essere stato grande: supponendo di dover ridurre di un quarto i 5 milioni di cui sopra, per tenere conto delle intestazioni ripetute allo stesso nome in distretti diversi e non eliminate, rimarrebbero sempre 3.750.000 proprietari maschi maggiorenni, numero che supera di un milione quello dato dal censimento del 1881.

Una seconda prova dei gravi errori del censimento a questo riguardo si ha pensando che su 145.000 successioni annue in Italia, 127.000 circa comprendono beni immobili e di queste la maggior parte (forse 100.000) sono di defunti maschi maggiorenni. Prendendo il solito moltiplicatore 36 (anni di durata di una generazione), si dovrebbero avere in Italia almeno 3.600.000 proprietari maschi maggiorenni viventi. Il censimento del 1901 ne ha indicato ancora un milione di meno. Esso poi come il precedente ha dato un numero di donne proprietarie piuttosto superiore al vero.

assente temporaneamente. Ora, la contropartita esatta delle due categorie di schede difficilmente si realizza. Nel 1881 si censirono 725 mila assenti temporanei dai diversi comuni, ma presenti nello Stato, ai quali avrebbero dovuto fare esatto riscontro i presenti occasionali nei diversi comuni. Invece questi presenti occasionali, dedotti gli stranieri di passaggio, non sommarono che a 483 mila. Differenza enorme, che pur sarebbe risultata maggiore, se invece di un riscontro *aritmetico*, in cui molti errori in opposto senso si compensano, si fosse fatto il riscontro *nominativo*.

Il censimento del 1901 è riuscito più esatto su questo punto. I presenti occasionali, dedotti gli stranieri di passaggio, risultarono in numero di 705 mila, contro 750 mila dichiarati assenti temporaneamente dalla famiglia, ma presenti nel Regno. La relativa precisione dei totali è però compatibile con errori rilevanti nei parziali. Ad ogni modo sta il fatto che vi sono assenti, i quali considerano abituale la propria dimora nel luogo in cui sono censiti, mentre le loro famiglie li denunziano come assenti temporanei — e vi sono censiti che dichiarano la dimora occasionale in un comune, mentre le famiglie li considerano assenti permanentemente e perciò si astengono dal riempire per essi una scheda. In ambo i casi, la contropartita vien meno, sebbene possa darsi compensazione di contrari errori. Inoltre se l'assente è persona che vive da sola, se è capo-famiglia di se stesso, chi ne certifica l'assenza? Vi sarà una scheda, quella originale, che lo indica presente occasionalmente in un dato comune, ma mancherà quella che lo indichi assente dal comune di abituale dimora.

Miglior partito sarebbe stato dunque *censire i soli presenti*, richiedendo da essi la semplice indicazione del comune di abituale dimora. Una domanda invece di tre, quindi impossibile l'equivocare; poche righe di istruzioni dovevan bastare per determinare la condizione speciale dei girovaghi, dei viaggiatori, degli impiegati in imminenza di trasloco, ecc. Si sarebbe fatto, è vero, un censimento di soli presenti; ma la popolazione residente o legale poteva, dacchè il legislatore lo aveva voluto, risultare egualmente da una facile operazione di spoglio: separando, comune per comune, le schede dei censiti che avessero indicato come luogo di dimora abituale un comune diverso da quello di presenza, e inviando queste stesse schede (o una copia delle medesime) ai comuni di residenza. Ogni comune sarebbe così venuto in possesso delle schede originali dei suoi assenti temporanei (purchè presenti nel Regno), unico mezzo logico per arrivare a quel riordinamento delle anagrafi municipali, che era tra i fini del legislatore. Così la popolazione residente, che si assume come base per l'applicazione delle leggi, si sarebbe ottenuta dalla semplice addizione dei presenti con dimora abituale nel comune e degli assenti, le cui schede originali fossero al comune stesso pervenute pel tramite della Direzione generale della Statistica.

Nè si obbietti che all'uopo sarebbe occorso per ogni comune tenere aperta una specie di contabilità con ognuno degli altri 8200 comuni del Regno! Anche prendendo un caso, punto favorevole alla nostra tesi, quale quello di Milano, non è supponibile che i 23 mila presenti occasionali di Milano provengano da tutti gli 8200 comuni del Regno; tutto al più si avrà la rappresentanza di 500 o 600 comuni. Così non è supponibile che i 15 mila assenti temporanei da Milano si trovino sparsi per tutti i comuni del Regno. In breve, il problema tecnico si sarebbe limitato per la Direzione generale della Statistica ad accantonare 700 mila schede di presenti occasionali, ripartirle secondo l'ordine alfabetico dei comuni di loro dimora abituale e rinviarle a questi comuni, i quali venendo così in possesso delle schede originali dei proprii assenti temporanei, potevano accingersi per davvero al riordinamento dei registri anagrafici.

LIBRO II.

LA STATISTICA COME FORMA DI INDUZIONE

CAPO PRIMO

I segni della presenza e del modo di operare delle cause.

Sommario: § 1. Preliminari. — § 2. Se i termini di una formola periodica corrispondano a distinte cause di variazioni. — § 3. La stessa questione per le formole algebriche intere. — § 4. Equivalenza o non-equivalenza dei gruppi di cause nelle seriazioni simmetriche e asimmetriche. — § 5. Le discontinuità delle serie e seriazioni come segni dell'intervento di cause speciali. — § 6. Momento in cui entra in azione o cessa d'agire una causa.

§ **1.** Nel risalire dai fatti alle cause loro o nel discendere da queste agli effetti il procedimento logico si compie alle volte così rapido e quasi per salti, da non lasciare a tutta prima chiara coscienza di stadii attraversati, intermedii fra il noto e l'ignoto. Quando di più persone, cui sia proposto un problema difficile, una o poche arrivano alla soluzione in questo modo, mentre molte non ci arrivano che seguendo una lunga catena di pensieri, noi ammiriamo nelle prime l'*intuito*, riconosciamo alle seconde l'attitudine logica. Non già però che l'intuito sia qualche cosa di opposto o di diverso dalla ordinaria facoltà di ragionare; ne è solo una manifestazione che si segnala per vivacità e prontezza. Gli stadii intermedii tra il noto e l'ignoto comprendono generalizzazioni induttive, ipotesi, percezioni di analogie, ecc., che in alcuni si affacciano quasi istantaneamente e vengono tosto utilizzate per ulteriori acquisti del pensiero, in altri invece si presentano in successione ad intervalli più distinti e spesso con ingombro di elementi inutili, che richiedono volta per volta un lavoro di eliminazione.

Ma per quanto abbreviati siano in molti casi i processi mentali, essi ammettono sempre una scomposizione in parti; il soggetto medesimo, in cui si compiono, per comunicare ad intelletti meno rapidi del suo il nuovo acquisto di verità, deve con un ritorno sul proprio lavoro logico distinguere per anelli la catena percorsa d'un tratto. La Logica, come scienza, non è che la teoria generale di questo percorso distinto per tappe.

Individuare le cause non è semplicemente dar loro un nome, nè il collocarle in una piuttosto che in altra delle categorie stabilite di fenomeni, nè il creare per esse, in certi casi, delle categorie nuove; ma vuol dire precisarne la direzione, l'intensità, il punto o momento d'applicazione, la durata di attività, l'interdipendenza con altri fattori, ecc. Naturalmente, a seconda delle difficoltà che presenta la soggetta materia, noi ci contentiamo di una determinazione più o meno completa.

Nell'ordine dei fatti collettivi, come già dicemmo nell'introduzione, il concetto di « causa » è più generico di quello accettato dalle scienze fisiche. Statisticamente si qualifica come *causa* anche un concorso di antecedenti necessari *e di circostanze d'azione ignota o dubbia non separabili da quelli.* In realtà la vera causa, nel significato delle scienze fisiche, si troverebbe al limite di una eliminazione perfetta degli elementi non necessari. Approssimarci a quel limite è quanto dobbiamo fare formando gruppi scelti, confrontando situazioni simili per molti aspetti, dissimili per qualcuno; e così via.

§ 2. I segni della presenza, della comparsa e scomparsa di cause o gruppi di cause speciali son resi visibili dai tracciati grafici delle serie e seriazioni statistiche. L'analisi matematica tende poi alla discriminazione dei fattori e alla valutazione del loro modo d'operare. E appunto il problema più arduo e insieme più interessante della Statistica matematica è di sapere se i singoli termini di sviluppo di una funzione algebrica intera o di una periodica possano considerarsi come espressione e misura di cause distinte o di gruppi distinti di cause. Per intenderci meglio procederemo, al solito, per via d'esempi.

La curva descritta da un proiettile lanciato nel vuoto può porsi sotto la forma: $y = bx + cx^2$, per modo che il termine bx stia a rappresentare la forza iniziale, che agisce in funzione semplice dei tempi, e il termine cx^2 la gravità, che agisce in ragion dei quadrati dei tempi. Le due cause, dalla cui combinazione deriva il movimento parabolico, avrebbero così la loro espressione e misura nei due distinti termini della formola. Senonchè qui la specie delle forze, e per una di esse (la gravità) anche la misura, s'intende anticipatamente conosciuta, conosciuta cioè per razionali principii o per precedenti induzioni; la scelta della formola non ha quindi niente di arbitrario.

Così pure, diremo collo Schiaparelli « allorchè si ha da determinare il corso di un pianeta dietro le osservazioni, già è noto a qual classe di curve e di velocità appartiene il suo movimento e di questo sappiamo scrivere tutte le particolarità in formole analitiche di certissima espressione, le quali altro d'indeterminato non involgono, che il valore numerico di un piccolo numero di coefficienti. Ogni osservazione dà allora una equazione tra quantità misurate e i coefficienti in questione e la somma perfezione dell'indagine consiste nel trarre dall'intera massa dei dati osservati quel sistema di parametri, che ha maggior probabilità

di accostarsi al vero. Il problema ammette una forma generale di soluzione adattabile a tutte le questioni — ed è noto che la sua trattazione costituisce il *metodo dei minimi quadrati* ».

Ben altro è il caso, quando senza nulla conoscere preventivamente delle cause in giuoco, adottiamo ad arbitrio una formola d'interpolazione e ricerchiamo poi se i singoli termini di essa rispecchiano l'azione di cause distinte. L'astronomo Carlini ce n'ha dato un elegante esempio a proposito della legge della *variazione diurna del barometro*. Questa variazione diurna può essere abbastanza bene rappresentata da una formola periodica coi termini in φ e 2φ, ossia dalla somma di due variazioni distinte, una avente il periodo di 24 ore e l'altra quello di 12. Tale formola, da noi già ricordata a pag. 176, nota 2, è:

$$y = \text{mm. } 751{,}793 + \text{mm. } 0{,}524 \text{ sen } (163^\circ{,}9' + \varphi) + \text{mm. } 0{,}227 \text{ sen } (129^\circ{,}48' + 2\varphi).$$

Il Carlini attribuì la prima variazione agli effetti molteplici del calore solare nell'atmosfera (*flusso fisico*, com'egli lo chiamava), che ha appunto un ciclo principale unico in 24 ore, e la seconda all'attrazione del sole sull'oceano atmosferico (*flusso dinamico*), che ha nella giornata due periodi di 12 ore ciascuno. La sua spiegazione non appaga tuttavia lo Schiaparelli e il Celoria, che oppongono argomenti, in cui non possiamo entrare; essi ritengono assai verosimile « che la separazione dei termini della formola periodica sia, come in generale è sempre riguardo ai fenomeni della meteorologia, un puro fatto analitico dipendente dalla forma arbitraria della funzione scelta come tipo, e non già l'espressione di un fatto fisico. È possibile, aggiungono, che la variazione totale del barometro sia la somma di due azioni distinte, come quelle supposte dal Carlini, ma non pare accettabile la conclusione che gli effetti di tali azioni si manifestino separatamente nei due primi termini di una formola empirica, mentre l'osservazione mostra l'esistenza di termini ulteriori non trascurabili ». Con queste ultime parole i due astronomi alludono ad un termine in 3φ, a un'onda tripla della variazione diurna, che è specialmente sensibile nei solstizii invernali e il cui periodo di 8 ore non si saprebbe a qual causa riferire, se non allo stesso flusso fisico o al flusso dinamico (1).

A noi pare che la questione veramente dovrebbesi invertire e presentarsi così: se per razionali principii o per antecedenti induzioni si ammette che l'attrazione solare sull'atmosfera, manifestantesi con due massimi e due minimi nella giornata, e il calore solare col suo ciclo

---

(1) Veggasi: *Sulle variazioni periodiche nel clima di Milano*. Memoria di G. V. SCHIAPARELLI e G. CELORIA, astronomi della Specola Reale di Brera, in Milano. Inoltre F. CARLINI, *Sulla legge delle variazioni orarie del barometro* (*Memoria per la Società italiana delle Scienze*, vol. XX, Modena 1828).

unico (1) entrino come fattori di variazioni diurne del barometro — ed altre cause paragonabili a queste due per importanza non si conoscano — è certo che l'interpolazione della formola periodica in termini di sen φ e di sen 2φ, rispecchierà in qualche modo quelle due azioni distinte. Incertezza resterà riguardo al valor relativo dei coefficienti, ma sarà tanto minore quanto più i residui del processo interpolatorio avranno carattere di perturbazioni accidentali o di errori d'osservazione. Sicchè i termini di una formola analitica possono essere segni di cause distinte; ma non lo sono effettivamente se non quando le cause immaginate si dimostrano per altra via capaci degli effetti loro attribuiti. La qual restrizione non toglie nulla al valore *suggestivo* (diciamo così) delle formole interpolatorie.

Analogo a quello del Carlini è in demografia il caso della variazione annua della *natalità legittima*. L'equazione della curva, pei dodici mesi da luglio a giugno, fatta eguale a 1000 la media mensile, è:

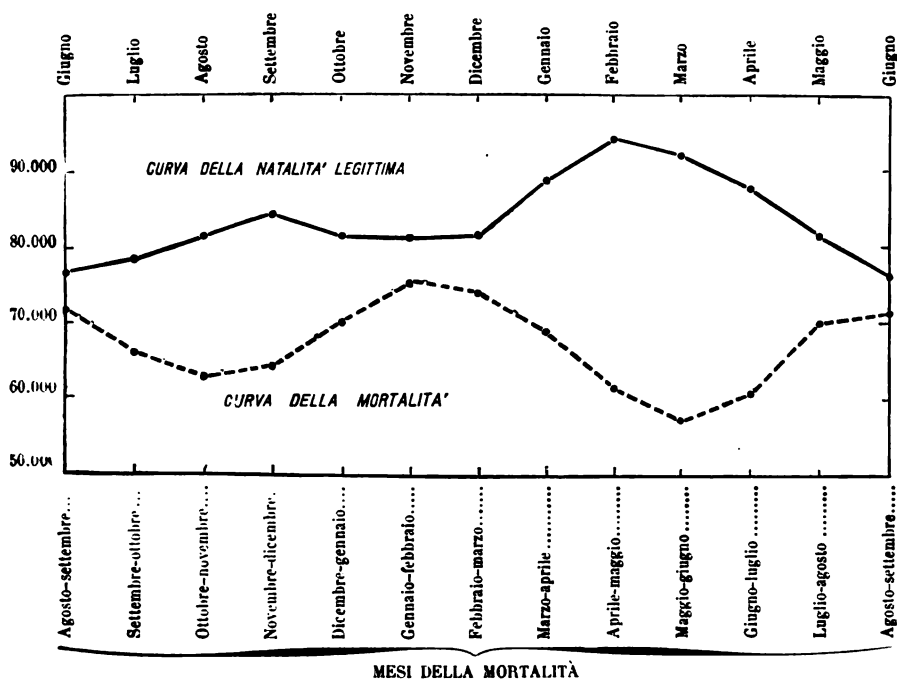
$$y = 1000 - 35{,}7 \text{ sen } \varphi - 57{,}8 \cos \varphi + 54{,}8 \text{ sen } 2\varphi - 9{,}3 \cos 2\varphi.$$

Anche qui abbiamo un'onda semplice (termini in φ), complicata con un'onda doppia (termini in 2φ). Ora tra i fattori immaginabili di variazioni della natalità legittima son certo da mettere in prima linea: *a*) la variazione annua dei *matrimoni*, a principiare dal decimo o dall'undecimo mese avanti le nascite; *b*) la variazione annua della *mortalità* a principiare dal nono o dal decimo mese avanti le nascite. Il primo fattore non ha bisogno di essere dimostrato. Quanto al secondo, è a ritenersi che nelle stagioni, in cui si aggravano la morbilità e la mortalità, presso molte famiglie colpite o anche solo minacciate d'un lutto domestico, i rapporti sessuali dei coniugi si facciano più rari; più rari quindi i concepimenti e a tempo debito le nascite. Il contrario, in corrispondenza delle epoche di morbilità e mortalità miti.

Nel corso dell'anno la nuzialità si presenta come un'onda unica, con due perturbazioni nel novembre-dicembre e nel febbraio-marzo, le quali non debbono avere un contraccolpo apprezzabile sulla curva delle nascite, in quanto i loro effetti diffusi anche sui mesi precedenti e susseguenti al decimo o all'undecimo dalle nozze avvenute, vengono in certo modo a perequarsi da sè stessi. La mortalità invece ci si presenta con due massimi (febbraio e agosto) e due minimi (maggio e ottobre):

(1) In realtà se il flusso fisico si deve ascrivere al calore solare, siamo in un caso analogo a quello della variazione diurna della temperatura, la cui equazione contiene non solo un termine in φ (onda semplice), ma anche uno non trascurabile in 2φ (onda doppia). Sicchè nella formola della variazione diurna barometrica, il termine in 2φ potrebbe derivare dal concorso di entrambe le azioni, dinamica e termica, del sole sull'atmosfera, e non dalla prima soltanto.

l'interpolazione della serie segnalerebbe qui come nella natalità un'onda semplice complicata con una doppia (1).

Nel diagramma che segue, la correlazione inversa tra natalità e mortalità vien bene in evidenza; congiungendo tra loro con una curva continua i due massimi del primo fenomeno e i due minimi del secondo, si avrebbe il probabile andamento della natalità nell'ipotesi che la mortalità presentasse un'onda unica e non duplice nel corso dell'anno.

Or dunque la seconda parte della formola della natalità (termini in 2φ) con tutta verosimiglianza dipende principalmente dall' onda doppia della mortalità; mentre la prima (termini in φ) ci fa pensare all'azione combinata dell'onda semplice della nuzialità e di quella semplice della stessa mortalità.

Con ciò non si esclude il concorso di altri elementi; per dirne uno, l'*emigrazione*, il cui andamento è pure a onda doppia (con un vertice principale in marzo-aprile e uno secondario in ottobre-novembre) tale da correggere le non rigorose coincidenze di tempo fra i massimi della

---

(1) Veggasi: *Gli aspetti arbitrarii dell'interpolazione delle serie statistiche* (*Giornale degli Economisti*, gennaio 1904). L'equazione della curva della mortalità, fatta eguale a 1000 la media mensile, è a cominciare dal settembre:

$$y = 1000 + 54{,}9 \text{ sen } \varphi - 19{,}0 \cos \varphi$$
$$- 100{,}9 \text{ sen } 2\varphi + 50{,}7 \cos 2\varphi.$$

natalità e i minimi della mortalità. Per quanto i coniugati figurino in minoranza tra gli emigranti, devesi ammettere che per essi un certo contributo venga sottratto alle nascite in epoche corrispondenti, a nove o dieci mesi d'intervallo, a quelle di forte emigrazione.

Naturalmente, se uno stesso termine della formola dipende da due o più cause in combinazione, rimane la difficoltà di assegnare a queste il valore relativo; il che richiede un vero e proprio lavoro d'induzione. Bisognerà perciò formare gruppi scelti colle annate o colle provincie, in cui la mortalità è stata eccezionalmente forte o eccezionalmente mite, ovvero ha avuto il suo massimo principale nell'estate anzichè nell'inverno, ecc. Inoltre, essendo verosimile che la mortalità eserciti una diversa influenza sul *moral restraint* dei coniugi secondo le classi d'età che essa colpisce maggiormente nelle varie stagioni, converrà non considerarla nel suo complesso, ma scendere a qualche specificazione. Così dicasi dei gruppi scelti da formare in riguardo alla nuzialità e alla emigrazione. Dal confronto delle diverse situazioni si ricaveranno, con metodi analoghi a quelli dell'induzione sperimentale, i valori relativi cercati.

Concludendo, a noi non sembra abbastanza giustificata l'idea pessimista dello Schiaparelli e del Celoria, che la rappresentazione dei fenomeni meteorologici, come dei demografici, ecc., per mezzo di formole analitiche « non ci avanzi neppure di un passo verso la cognizione della vera legge di tali fenomeni ». Al contrario, troviamo ben suggestivi questi processi, coi quali si tenta di risalire, nella dinamica delle serie, dalla risultante alle componenti e che invitano a ricercare dietro il velo dei simboli le cause specifiche dei fatti. Ben inteso che il còmpito dello statistico non deve fermarsi al primo aspetto formale, che le cose gli presentano all'analisi matematica; affinchè non i fatti si pieghino alle formole, ma le formole si adattino ai fatti, occorre l'esercizio sicuro della induzione e della deduzione.

§ **3.** La questione trattata, che riesce relativamente semplice nel caso di serie periodiche o a ciclo chiuso, si fa difficile nel caso di serie aperte o continue, analiticamente esprimibili per via di funzioni algebriche intere. Sia una serie statistica rappresentata nelle sue maggiori variazioni della formola:

$$y = ax^3 + bx^2 + cx + d.$$

Ai singoli termini in x, in x^2, in x^3 corrisponderanno cause distinte? Ci sarà una causa che agisce in funzione semplice dei tempi, un'altra che agisce secondo i quadrati dei tempi e una terza secondo i cubi?

Nel mondo fisico queste forme d'azione sono tutt'altro che rare; ma così non può dirsi dei fatti demografici, economici, ecc. Alcuna volta, trattandosi, ad esempio, di fenomeni che da un centro di diffusione si propagano gradatamente nel territorio o da un ceto scelto di persone

si propagano per imitazione a classi inferiori e più numerose, può risultare un andamento parabolico della serie, ma in generale la regola dello sviluppo secondo le potenze intere della variabile (tempi) non si troverà verificata che con una certa latitudine. D'altronde un termine in x^2 o in x^3 non ha bisogno d'essere spiegato da *una* causa operante giusto secondo la norma dei quadrati o dei cubi dei tempi, ma può nascere dalle più svariate combinazioni di cause operanti altrimenti. Così il *prodotto* di due funzioni lineari — che è quanto dire di due progressioni aritmetiche — deve dare un termine in x^2. Se in un paese la quantità fisica prodotta di una merce variasse per un certo tempo come la funzione $y = a + bx$, e il prezzo unitario come la funzione $y' = c + dx$, la curva del valore totale sarebbe parabolica di secondo grado. Infatti:

$$(a + bx)\,(c + dx) = ac + (bc + ad)\,x + bd\,x^2.$$

Un termine in x^2 può risultare anche dall'addizione di cause costanti e di cause intermittenti con azione uniforme nell'intervallo. Così il tratto di serie esattamente parabolico

20 . 26 . 29 . 29 . 26 . 20

potrebbe dissimulare un concorso di antecedenti di questo genere:

Primo antecedente	20	20	20	20	20	20
Secondo »	—	6	6	6	6	—
Terzo »	—	—	3	3	—	—
Somma . . .	20	26	29	29	26	20

Veniamo ad un caso concreto. La *natalità illegittima* in Italia, dopo introdotto il matrimonio civile è rapidamente cresciuta per fatto delle unioni che in alcune provincie continuarono a celebrarsi col solo vincolo religioso. Considerando il periodo 1872-1900, la curva delle nascite illegittime si vede ascendere fino al 1885 e indi piegare in basso. L'equazione della interpolatrice (fissata l'*origine* al 1886) è:

$$y = 81.970 - 108{,}5\,x - 87{,}8\,x^2.$$

Ora il ramo ascendente di questa parabola può essere spiegato da un numero non crescente, ma *costante*, di unioni irregolari; invece il ramo discendente presuppone un numero non più costante, ma *decrescente*, di quelle, solo che si ammetta che la fecondità delle coppie pseudo-coniugali diminuisce col passare del tempo in progressione aritmetica, la qual cosa è abbastanza verosimile, se si tien conto oltre che delle conseguenze naturali dell'età, anche dell'effetto che deriva dalle regolarizzazioni di molte convivenze per susseguente matrimonio civile.

Ciò ammesso, il seguente schema dimostra che alla lunga, rimanendo costante il numero annuo delle unioni contratte col solo vincolo religioso, la linea delle nascite illegittime di parabolica si convertirebbe

in una retta parallela all'asse dei tempi; perchè quella conservi il suo carattere parabolico iniziale e dia luogo, come infatti ha dato, ad un ramo discendente, bisogna ammettere che, *coeteris paribus*, il numero delle unioni irregolari, di costante che era, abbia da un certo momento cominciato e poi seguitato a diminuire. Ecco lo schema:

Unioni irregolari	10	10	10	10	10	10	10	10 .....
Fecondità della 1ª schiera	5	4	3	2	1	—	—	— .....
» » 2ª »	—	5	4	3	2	1	—	— .....
» » 3ª »	—	—	5	4	3	2	1	— .....
» » 4ª »	—	—	—	5	4	3	2	1 .....
» » 5ª »	—	—	—	—	5	4	3	2 .....
» » 6ª »	—	—	—	—	—	5	4	3 .....
» » 7ª »	—	—	—	—	—	—	5	4 .....
» » 8ª »	—	—	—	—	—	—	—	5 .....
Somma . . .	5	9	12	14	15	15	15	15 .....

Fino alla sesta unità di tempo i numeri della somma formano una serie parabolica di secondo grado, che si convertirebbe poi in una serie costante, se da quel momento le unioni irregolari non cominciassero a diminuire. Per analogia concluderemo, *in via di prima approssimazione*, che in Italia si dovette avere fino al 1885 un numero piuttosto costante di unioni sancite col solo rito religioso e da quell'anno in poi un numero decrescente.

Anche qui dunque non abbiamo avuto necessità di far corrispondere il termine in x^2 ad una causa operante in ragion dei quadrati dei tempi, ma ad una particolare combinazione di cause operanti con altra norma. Ad ogni modo resta acquisito che i termini della formola analitica, debitamente interpretati, possono apprenderci qualche cosa sul modo d'azione degli antecedenti necessari di un fenomeno.

Di questo avviso è il Pareto. Egli dice: « Quando si applica la formola d'interpolazione:

$$y = A + B\psi_{I} + C\psi_{II} + \ldots$$

si osserva in generale che le curve semplici, che si ottengono successivamente, non vanno avvicinandosi d'un passo uniforme alla curva reale; la *precisione* comincia dall'aumentare rapidamente: in seguito c'è uno stadio in cui aumenta con lentezza; poi daccapo con rapidità, e così via. Questi stadii o intervalli, nei quali la precisione cresce con lentezza, separano i grandi gruppi di sinuosità; in altre parole, essi separano dei *gruppi d'influenze* via via più speciali, che si esercitano sul fenomeno » (1).

---

(1) Veggasi: *Quelques exemples d'application*, ecc., pag. 5.

§ 4. Nel caso di variazioni degli organi di piante e di animali, che dànno luogo a seriazioni simmetriche attorno alla *media* o asimmetriche attorno alla *norma*, vedemmo accettata dai biologi l'ipotesi che la simmetria dipenda dal concorso di *due gruppi di ignote, ma in complesso equivalenti cause elementari*, le quali partecipano alla formazione delle deviazioni del carattere considerato, in eccesso o in difetto sul suo valor medio; e che l'asimmetria dipenda invece dalla *non-equivalenza* dei due gruppi di cause. Poichè abbiam dato un esempio di seriazioni simmetriche (gli arruolati italiani distinti secondo l'indice cefalico), vogliamo qui darne qualcuno di seriazioni asimmetriche, togliendole dal Davenport e Bullard, dal Duncker e dal Weldon.

Ghiandole mülleriane del maiale		Spine della pinna dorsale nell'*Acerina cernua*		Rapporto tra la parte frontale e la lunghezza del corpo nel *Palaemonetes varians*	
Numero delle ghiandole	Casi osservati	Numero delle spine	Casi osservati	Rapporto p. °/oo	Casi osservati
0	15	11	1	580-591	9
1	209	12	2	592-603	19
2	365	13	189	604-615	52
3	482	14	1234	616-627	96
4	414	15	454	628-639	176
5	277	16	20	640-651	244
6	134	—	—	652-663	256
7	72	—	—	664-675	114
8	22	—	—	676-687	33
9	8	—	—	Oltre 687	1
10	2	—	—	. .	—
	2000		1900		1000

Consideriamo la prima seriazione e applichiamole la formola del Pearson (veggasi pag. 233, nota 2):

$$y = y_0 \left(1 + \frac{x}{a_1}\right)^{v a_1} \left(1 - \frac{x}{a_2}\right)^{v a_2}$$

dove y_0 è la classe in cui è compresa l'ordinata massima; a_1 il segmento dell'asse delle ascisse fino al quale si estende il ramo di sinistra della curva interpolatrice; a_2 il segmento fino al quale si estende il ramo di destra; v è una costante. L'equazione della interpolatrice è:

$$y = 473{,}9 \left(1 + \frac{x}{4{,}29}\right)^{4{,}84} 1 - \frac{x}{15{,}60}\Big)^{17{,}62}.$$

Le costanti caratteristiche sono a_1 e a_2, le quali misurano l'estensione dei due rami della curva asimmetrica. I loro valori, 4,29 e 15,60,

ci dànno un'idea della portata relativa dei due gruppi di cause, producenti le deviazioni in più o in meno della norma. L'individuazione delle cause si limita, è vero, alla separazione ideale di due gruppi contrari di antecedenti e alla valutazione *in globo* della loro relativa efficacia; ma è sempre un guadagno non disprezzabile delle nostre cognizioni in argomento.

§ **5.** Le discontinuità improvvise nelle serie e seriazioni sono generalmente segni della comparsa o scomparsa di cause speciali. È ovvio che se i fattori, i quali hanno dato una certa forma iniziale alla curva, rimanessero gli stessi per numero e specie, la curva seguiterebbe a svolgersi con carattere di perfetta continuità. Una soluzione di continuità non può dunque spiegarsi se non come effetto di qualche elemento nuovo, che si è aggiunto al concerto dei precedenti fattori, o di una perdita di elementi, che questo medesimo concerto ha subito d'un tratto. Acquisti e perdite di elementi sono in verità la condizione normale dei fenomeni statistici; in altri termini, le cause di questi formano un intreccio instabile, un sistema perennemente perturbato. Le corrispondenti discontinuità nelle serie e seriazioni o sono di grandezza compatibile coll'ipotesi di una probabilità costante del fenomeno, e allora noi le presumiamo prodotte da minutissime influenze, accidentali e inaccessibili all'analisi; o sono di grandezza mal conciliabile con quell'ipotesi, e allora le consideriamo segni di cause speciali, che la induzione statistica deve saper individuare. Il tener nota di questi segni diventa così un'operazione logica preliminare della induzione. Talvolta però essi non vengono in evidenza se non in seguito ad una elaborazione intensiva dei dati; ed è di questo caso che ci piace recare un esempio.

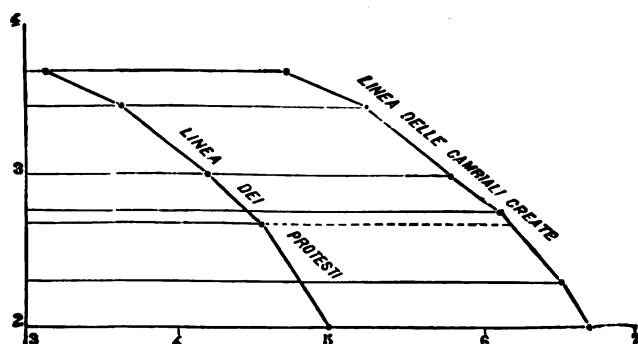
Le cambiali create in Italia nel periodo 1890-91 - 1895-96 e quelle cadute in protesto nel periodo 1891-96, formano due seriazioni difficilmente paragonabili nei dati originarii, non tanto per qualche divario che esse presentano nella scala di graduazione per valore, quanto per i motivi che indichiamo in nota (1).

---

(1) Veggasi, per le cambiali create, il *Bollettino di Statistica e Legislazione comparata*, anno III, fasc. 3°, pubblicato dalla Direzione generale del Demanio e Tasse sugli affari, e per i protesti le *Statistiche giudiziarie civili*, pubblicate dalla Direzione generale della Statistica. I dati delle prime, nel nostro prospetto, risultano da una speciale elaborazione fatta subire agli elementi statistici originarii, e, a partire dal limite di 600 lire, debbono considerarsi approssimativi. La Direzione generale del Demanio e Tasse sugli affari ha tratto la classificazione degli effetti cambiarii dalle vendite di carta bollata a tassa graduale; ma è noto che questa carta per gli effetti del valore di 1000 a 2000 lire, di 3 a 4 mila, di 7 a 8 mila, di 9 a 10 mila, serve anche per gli effetti *a scadenza maggiore di 6 mesi* del valore di 600 a 1000 lire di 1000 a 2000, di 2 a 3 mila, di 3 a 4 e di 4 a 5 mila, rispettivamente. Abbiamo quindi dovuto riportare al loro giusto posto di valore le cambiali a scadenza maggiore di 6 mesi,

Il confronto sui dati elaborati viene a stabilirsi così:

CAMBIALI CREATE		PROTESTI CAMBIARII	
Ammontare	Numero medio annuo	Ammontare	Numero medio annuo
Sotto 100 lire	2.992.379	Sotto 100 lire	24.409
Da 100 a 200 lire . .	1.576.385	Da 100 a 500 lire . .	60.784
» 200 a 600 » . .	2.128.259	» 500 a 1000 » . .	20.437
» 600 a 1000 » . .	657.307	» 1000 a 3000 » . .	11.270
» 1000 a 3000 » . .	471.650	» 3000 a 5000 » . .	2.836
» 3000 a 5000 » . .	125.446	Oltre 5000 lire. . . .	1.376
Oltre 5000 lire	54.248		
Totale . . .	8.005.674	*Totale* . . .	121.112

Integrate queste due seriazioni (calcolando il numero degli effetti da 100 lire in su, da 200 in su e così via) e rappresentate in diagramma a doppia scala logaritmica, dànno luogo alla seguente figura:

Rappresentazione a doppia scala logaritmica degli effetti cambiarii creati e di quelli caduti in protesto a partire dal limite minimo di 100 lire.

La differenza, che ad occhio si potrebbe apprezzare costante, tra le due curve dei protesti e delle cambiali emesse — almeno dal limite di 600 lire in poi — denota che *gli effetti di medio ammontare e quelli di alto ammontare hanno a un dipresso la stessa probabilità di cadere in protesto*. Infatti, se sono costanti le *differenze* tra due serie di logaritmi,

---

calcolandone il numero in ragione di 3,8 cambiali a lunga per 100 a breve scadenza, rapporto accettabile in base ad alcuni elementi di calcolo forniti dalla citata pubblicazione per gli effetti non superiori a 600 lire. Infine per conseguire una certa compensazione degli errori non potuti eliminare, si è proceduto per classi di valore piuttosto larghe.

costanti debbono essere i *rapporti* tra i numeri, cui quei logaritmi corrispondono. Noi siam dunque tratti a pensare che quel complesso di cause (congiunture sfavorevoli, errori di previsione, poca probità commerciale, ecc.), per cui si vien meno alla promessa di pagamento dei debiti alla scadenza, agisce con forza press'a poco eguale presso le persone, che fanno affari di mediocre importanza e presso quelle che fanno grossi affari. Tuttavia, la natura stessa di questa rappresentazione logaritmica (da cui è esclusa la categoria degli effetti inferiori a 100 lire, per essere ignoto il loro limite minimo di valore) e la piccolezza della scala adottata, lasciano incerto se l'identica conclusione possa estendersi ai debiti di piccolissima entità. Per chiarire la cosa convien ridurre le due seriazioni allo stesso tipo di classificazione per valore, ciò che si ottiene con un facile procedimento interpolatorio, diretto a distinguere anche nelle cambiali emesse la categoria da 100 a 500 lire, e quella da 500 a 1000 (1). Il risultato è il seguente:

Ammontare	Cambiali create — Numero medio annuo	Protesti — Numero medio annuo	Rapporto dei protesti alle cambiali create
Sotto 100 lire	2.992.379	24.409	0,82 p. 100
Da 100 a 500 lire . .	3.347.555	60.784	1,81 »
» 500 a 1000 » . .	1.014.596	20.437	2,01 »
» 1000 a 3000 » . .	471.650	11.270	2,39 »
» 3000 a 5000 » . .	125.446	2.836	2,26 »
Oltre 5000 lire	54.248	1.376	2,54 »
Totale . . .	8.005.674	121.112	

La piccolezza del rapporto per gli effetti inferiori a 100 lire è troppo eccezionale per non indurre in sospetto di una causa speciale. Non v'è ragione perchè, se le altre cause generali agiscono uniformemente, o quasi, sui titoli da 100 lire in su, non agiscano allo stesso modo su quelli da 100 in giù. Quasi si sarebbe indotti a pensare che per molti titoli di minimo importo, non pagati alla scadenza, si tralasci di levare il protesto, lasciandosi loro la semplice efficacia di obbligazioni civili; mentre ciò non succede certo in caso di somme considerevoli. Se la supposizione trovasse conferma da parte dei pratici, ben si potrebbe dire che il segno di una causa specialissima esisteva ed era riconoscibile nella discontinuità dei primi termini della serie di percentuali.

(1) Il numero delle cambiali emesse del valore da 500 lire in su fu calcolato per interpolazione lineare sui logaritmi del numero delle cambiali di ammontare superiore a 200 lire e di quello delle cambiali di ammontare superiore a 600. Indi, procedendo per differenze, si ottennero le due categorie distinte da 100 a 500 e da 500 a 1000.

§ 6. Quello che in meccanica si direbbe il *punto d'applicazione* di una forza, in statistica è il *momento* in cui entra in azione o cessa di agire una causa o un gruppo di cause. Alle volte l'entrata in azione o l'uscita avvengono per gradi insensibili, che solo un delicato processo di confronti può indicare, sempre però a condizione che le statistiche siano abbastanza specializzate. Nella previsione di un cattivo raccolto o di un inasprimento di dazi si anticipano gli approvvigionamenti di derrate dall'estero; qual è il momento iniziale di attività della speculazione? Il cambio in Italia si era trasformato gradatamente in aggio, ancor prima del 1894, ancor prima cioè del ripristino ufficiale del corso forzoso; da qual momento preciso cominciò quella trasformazione? L'analisi completa delle cause comprende pure questo genere di determinazioni, a cui si arriva con metodo semiologico. Ma la varietà dei casi è tanta, che ogni buon tentativo di semiotica generale statistica sarebbe bene accetto.

Nelle serie periodiche le cause, da cui dipendono i massimi e i minimi, possono ripresentarsi ad intervalli non regolari o ad intervalli regolarmente variabili. In certo modo il punto d'applicazione della forza, che entra in gioco ad ogni periodo, non è fisso, ma è soggetto a spostamenti. Questa mobilità è ancora uno dei caratteri che per segni estrinseci dati dalle rappresentazioni geometriche o dalle analitiche si debbono precisare. Noi osserviamo, ad esempio, che in certe annate la maggior frequenza di nozze si ha nel gennaio e nell'aprile; in altre invece questi due mesi cedono buona parte di attività, l'uno al febbraio, l'altro al marzo; senza dire poi dei casi intermedii. Causa di ciò è la mobilità della Pasqua, che può cadere tra il 22 marzo e il 25 aprile; la qual causa, se già non si conoscesse per induzione comune, si potrebbe accertare col metodo delle variazioni concomitanti, del quale parleremo tra breve.

La decomposizione dei fenomeni di massa in gruppi scelti può rivelare segni, che nella massa indistinta sfuggirebbero all'occhio più penetrante. Fenomeni apparentemente non periodici risultano da compensazioni di periodicità opposte dei loro gruppi scelti; movimenti generali in una direzione determinata risultano spesso da movimenti particolari in direzioni anche molto diverse. Come altrove si è detto, le leggi statistiche non sono che interferenze complesse di leggi esatte, ma presentantisi con tali caratteri di uniformità da *parere* esse stesse delle leggi esatte.

CAPO SECONDO

I metodi d'induzione sperimentale e loro applicabilità ai fatti collettivi.

Sommario: § 1. Metodo di *concordanza.* — § 2. Metodo di *differenza.* — § 3. Metodo dei *residui.* — § 4. Esempio di combinazione dei tre metodi di eliminazione. — § 5. Metodo delle *variazioni concomitanti.* — § 6. Enumerazione degli antecedenti.

§ **1.** Lo Stuart Mill nel suo *System of Logic* ha ampiamente illustrati i quattro metodi di induzione sperimentale, detti *di concordanza, di differenza, dei residui* e *delle variazioni concomitanti,* ma ne contesta l'applicabilità alle scienze politiche e sociali.

Il metodo « di concordanza » si enuncia così: *Se due o più casi, in cui il fenomeno accade, presentano tutte le circostanze diverse, eccetto una, quest'unica circostanza in cui tutti i casi concordano, è causa o parte necessaria della causa del fenomeno.*

Questa regola di logica fondasi sull'assioma: ogni antecedente, la cui mancanza lascia sussistere l'effetto, non è necessario alla produzione dell'effetto, e quindi non fa parte della causa.

Supponiamo, dice il Bain, che A rappresenti una causa ed *a* un effetto. In natura è raro che A sia seguito solamente da *a*. Quando questa condizione estremamente semplice si avvera, l'induzione sperimentale non è necessaria. Invece è senza confronto più comune il fatto che A sia accompagnato da circostanze come B e C; e parimente *a* da circostanze come *b* e *c*. Se queste colleganze fossero invariabili, noi non avremmo alcun modo d'applicare il metodo di eliminazione. Sì l'abbiamo, quando quelle si presentano sotto diverse forme.

Per esempio:

A, B, C producono *a, b, c*
A, B, D » *a, b, d*
A, C, E » *a, c, e.*

Se si guarda solo alla prima serie, parrebbe che l'effetto *a* potesse provenire così da A, come da B, come da C. Ma se consideriamo la seconda serie, vediamo sussistere l'effetto *a*, sebbene manchi C nel gruppo degli antecedenti. Dunque C non è causa o parte necessaria della causa di *a*. Se guardiamo anche alla terza serie, troviamo sussistere ancora *a*, pur mancando B nel gruppo degli antecedenti. Anche B dunque non è parte necessaria della causa di *a*. Il solo antecedente comune alle tre serie è A; dovunque *a* è presente come conseguente, A lo è come antecedente. Pertanto se noi acquistiamo la certezza, con

l'esame di ulteriori serie, che ogni altra circostanza all'infuori di A può essere così eliminata, senza che venga meno l'effetto *a*, noi concluderemo che A è la causa, la condizione o l'antecedente invariabile di *a*.

Lo Stuart Mill contesta l'applicabilità di questo primo metodo di eliminazione alle scienze sociali. Il suo pensiero può riassumersi così: Il fisico e il chimico possono, variando a piacere le condizioni degli sperimenti, arrivare alla convinzione che una data circostanza è proprio l'unica comune a tutti i casi, mentre le altre son tutte diverse; invece l'osservatore dei fenomeni sociali, che per loro natura sono estremamente complessi e non si lasciano riprodurre *ad libitum* in condizioni opportunamente prestabilite, ha sempre ragion di dubitare che vi siano altri antecedenti comuni, oltre quello che ha fermato prima la sua attenzione e che, pur essendo presente in tutti i casi osservati, potrebbe non costituire ancora la condizione *sine qua non* del verificarsi del fenomeno.

Noi crediamo invece che almeno nell'ordine dei fatti collettivi, statisticamente accertabili, il metodo di concordanza sia applicabile, pur riconoscendo che qui l'induzione, grazie appunto alla maggiore complessità delle cause, conduce a risultati non certi, ma solo probabili. Tuttavia la probabilità o verosimiglianza può essere accresciuta dalla conferma dei risultati ottenuta coll'applicazione degli altri metodi, di *differenza*, dei *residui* e delle *variazioni concomitanti*. È questo un punto di capitale importanza metodologica. Ciascuno dei metodi nominati, isolatamente adoperato, ammette conclusioni che hanno solo un certo grado di probabilità; ma il concorso e il riscontro di tutti quattro i metodi può accrescere questa probabilità sino a convertirla in certezza pratica.

Il Pareto ci ha pel primo rivelato la singolare somiglianza nella ripartizione degli individui per redditi totali, in paesi e tempi diversi (V. addietro, pag. 187-188). Popoli agricoli e manifatturieri, protezionisti e libero-scambisti, a regime semi-feudale e di libera proprietà, cattolici e protestanti, progrediti ed arretrati, presentano all'incirca la stessa curva di ripartizione. Fra tante circostanze diverse, l'unica che al Pareto parve comune ai diversi paesi e tempi, sarebbe il modo col quale le attitudini utili all'acquisto della ricchezza si trovano distribuite nelle masse sociali. In altre parole, egli ritiene che la somiglianza formale nella ripartizione della ricchezza abbia radice in qualche cosa di inerente alla natura umana, nel fatto stesso che le qualità o doti fisiche, intellettuali e morali sono tanto più rare, quanto più eccellenti e si trovano dosate, per dir così, colla stessa norma nei grandi complessi di individui. Egli ne trae la conseguenza che, in una ipotetica società collettivista, la ripartizione dei beni non potrebbe differire di molto da quella che oggi è; i più abili ed intelligenti sapranno sempre assicurarsi la quota migliore e maggiore.

Il ragionamento del citato scrittore è di *induzione per concordanza*. I casi osservati presentano tutte le circostanze diverse, eccetto una, in cui tutti concordano; quest'una — che per altro è una circostanza piuttosto intuita che osservata — dev'essere la causa o parte necessaria della causa del fenomeno.

Il Pareto si arresta qui, mentre importerebbe pure vedere se la causa del fenomeno sta *tutta* o sta *solo in parte* nella circostanza accennata; ossia importerebbe vedere se non vi è proprio alcun altro antecedente comune ai casi osservati, per cui si possa spiegare la forma caratteristica costante della ripartizione dei redditi presso i diversi popoli e in diversi tempi. L'indagine è tanto più giustificata, in quanto che la ripartizione degli individui per doti fisiche, intellettuali, ecc. sembra seguire la legge degli errori accidentali; quindi se essa fosse l'unica causa della ripartizione della ricchezza, la curva di questa, anzichè *iperbolica*, apparirebbe *binomiale*. Ora quale può essere altro antecedente comune dei casi osservati? Può essere il *regime successorio dei beni*, che esiste infatti con poche differenze presso i diversi paesi osservati; e ad esso devesi soprattutto imputare se molti individui dalle attitudini inferiori riescono a mantenersi ad un gradino elevato nella scala della ricchezza ed altri dalle attitudini superiori non riescono, salvo in casi eccezionali, ad elevarsi sopra un'aurea media di reddito. La curva iperbolica dei redditi totali risulterebbe così da una deformazione della curva binomiale, dovuta all'elemento perturbatore della trasmissione ereditaria dei beni. Quindi ancora, in una ipotetica società collettivista che sopprimesse il regime successorio, la ripartizione degli individui per ricchezza non conserverebbe la forma attuale, ma più verosimilmente seguirebbe la nota legge degli errori.

§ **2.** Il metodo *di differenza* è pure un metodo di eliminazione. Si enuncia così: *Se i casi in cui il fenomeno si verifica, e quelli in cui non si verifica, hanno tutte le circostanze comuni, eccetto una, la quale si riscontra solo nei primi, la circostanza per la quale si differenziano le due serie di casi, è causa o parte necessaria della causa del fenomeno.*

A tale metodo si ricorre spesso nelle scienze fisiche. Senonchè in queste la natura dei fenomeni è abbastanza semplice da permetterci di acquistare la convinzione che una data circostanza è proprio l'unica, che stabilisce il divario tra la serie dei casi, in cui il fenomeno si produce e la serie di quelli in cui non si produce. In altri campi le cose procedono altrimenti. Il Mill fa questo esempio: sia da decidere se giovi più ad un paese il libero scambio o il protezionismo. Se due nazioni, egli dice, esistessero, in tutto simili (nelle condizioni favorevoli o sfavorevoli del territorio, nelle qualità fisiche o morali del popolo, nel reggimento politico, ecc.) *fuorchè nella politica commerciale*, e si trovasse che la nazione libero-scambista è più ricca e

la protezionista più povera, si avrebbe allora la prova sicura che il libero scambio promuove lo sviluppo della ricchezza nazionale e il protezionismo l'ostacola. Ma la supposizione, dalla quale si parte, è assurda anche in astratto. Due paesi che si somigliassero in tutti i caratteri fondamentali, si somiglierebbero necessariamente anche nella politica commerciale, che è una semplice derivazione o filiazione di quei caratteri. Se i due popoli seguono una opposta politica riguardo al commercio estero, segno è che diversificano per altre circostanze non notate e difficilmente accertabili nell'intreccio assai complesso degli antecedenti. Il problema è insolubile, non per difetto di elementi di soluzione, ma per eccesso. Nelle discipline morali quindi il metodo di differenza pare allo Stuart Mill inapplicabile.

Inapplicabile si, rispondiamo, se si pretende per esso di arrivare a conclusioni rigorose; applicabile se ci contentiamo di conclusioni approssimative, la cui probabilità possa essere accresciuta dall'accordo e controllo dei risultati conseguito con altri metodi. Tutti i logici, compreso il Mill, riconoscono il vantaggio di combinare i due metodi di concordanza e di differenza; nelle scienze morali, in ragione appunto della maggior complessità dei fenomeni, convien far di più e combinare i detti metodi con quello delle variazioni concomitanti. Inoltre, senza andar in traccia della unica circostanza comune o diversa a due serie di casi, ci si deve star paghi di riconoscere a vari segni la circostanza *principale*, trattando le altre di azione ignota o dubbia come elementi perturbatori.

Sia, ad esempio, da ricercare qual'è la circostanza principale e comune ai diversi casi, fra le molte d'azione ignota o dubbia o indifferente, per cui si spiega nelle città la presenza di un numero eccezionale di maschi, soprattutto celibi, nelle età dai 20 ai 30 anni — e la circostanza principale e diversa, fra le molte comuni o indeterminate, per cui si spiegherebbe il fenomeno opposto nelle piccole borgate e nei villaggi. Una diligente enumerazione delle caratteristiche delle varie città metterebbe in rilievo molte circostanze diverse (situazione, clima, specialità d'industrie, grado di ricchezza, ecc.); ne metterebbe in rilievo anche molte di comuni, ma che per altra via si possono dimostrare estranee o indifferenti all'effetto da spiegare; l'unica, o se non unica, principale, circostanza veramente comune e decisiva del fenomeno in questione è l'esistenza di uffici pubblici, guarnigioni militari, istituti di istruzione, industrie manifattrici, che reclutano appunto i loro elementi tra i giovani maschi dai 20 ai 30 anni, traendoli in buona parte dai centri minori e dalle campagne. Controlliamo il risultato coll'indagine di ciò che hanno di comune i centri minori e le campagne colle città e di ciò che hanno di diverso, tenendo presente il fatto che nelle prime si nota il fenomeno opposto, cioè una deficienza dei maschi dai 20 ai 30 anni in confronto delle femmine. Il rapporto dei sessi nelle nascite è press'a poco il medesimo nelle

campagne e nelle città; la mortalità dalla nascita ai 30 anni colpisce i due sessi in proporzioni simili, e così via; l'unica o principale circostanza veramente diversa e decisiva agli effetti di una alterazione del rapporto tra i due sessi nelle classi d'età dai 20 ai 30 è l'assenza nelle borgate e nei villaggi, e la presenza nei centri maggiori, di uffici pubblici, guarnigioni militari, istituti d'istruzione secondaria e superiore, industrie, ecc. La probabilità della conclusione, che se ne induce, può ulteriormente essere rinforzata col metodo delle *variazioni concomitanti*, osservando quel che si è verificato nelle piccole borgate, le quali nel decorso di un secolo sono divenute centri considerevoli di popolazione.

§ 3. Il metodo detto dei *residui* si enuncia così: *Togliendo da un fenomeno la parte, che per precedenti induzioni sappiamo essere l'effetto di certi antecedenti conosciuti, il residuo si dirà effetto degli antecedenti da determinare.*

Questo, meglio che un metodo d'induzione, è una forma di deduzione, di cui spesso si valgono le scienze sperimentali. In statistica pure l'applichiamo, sempre avvertendo che le conclusioni avranno una probabilità più o men grande, non la certezza. Infatti, per l'indole stessa dei fenomeni collettivi, quella parte dei medesimi, che per altre induzioni si attribuisce a certi antecedenti conosciuti, non è mai una quantità rigorosamente determinabile, ma solo approssimativamente; quindi anche il residuo sarà una quantità approssimativa.

Sia data la serie del consumo del *sale commestibile* in Italia (1):

Esercizi finanziari	Consumo di sale commestibile in quintali	Esercizi finanziari	Consumo di sale commestibile in quintali	Esercizi finanziari	Consumo di sale commestibile in quintali
1880-1881	1.460.287			1886-1887	1.666.904
1881-1882	1.473.051			1887-1888	1.673.253
1882-1883	1.491.810	1885-1886	1.601.781	1888-1889	1.702.735
1883-1884	1.512.611			1889-1890	1.728.461
1884-1885	1.544.790			1890-1891	1.745.581

La causa dell'aumento notevole del consumo è nota; fu il ribasso di 20 lire su 55 sul prezzo della derrata per quintale, ribasso stabilito per legge a partire dal 1° gennaio 1886. Il processo di individuazione delle cause non si limita alla indicazione del nome e del momento

(1) I dati dei primi tre esercizi risultano dalla semisomma di quelli degli anni solari 1880 e 1881, 1881 e 1882, 1882 e 1883. Il dato del 1883-1884 risulta dall'addizione di metà del consumo del 1883 e dal consumo del primo semestre 1884. I dati rimanenti si riferiscono ai veri e proprii esercizi finanziarii, che vanno dal 1° luglio d'ogni anno al 30 giugno successivo.

della loro comparsa, bensì comprende la determinazione dell'intensità e direzione dell'effetto. Il metodo dei residui può avere qui una interessante applicazione.

Interpoliamo una retta nella serie che va dal 1880-1881 al 1884-1885 e un'altra retta nella serie che va dal 1886-1887 al 1890-1891, fissata per entrambe l'origine all'anno di mezzo; lasciamo fuori conto l'esercizio 1885-1886, che fu per metà sotto un regime di tariffa, per metà sotto l'altro. Le equazioni delle rette sono rispettivamente:

$$y = 1.496.510 + 20.856,6\,x$$
$$y = 1.703.387 + 21.256,2\,x.$$

Il 1886-1887, primo anno in cui ebbe pieno effetto la riduzione del prezzo, sarebbe il 4° a partire dal 1882-1883, fissato come origine delle x nella prima equazione, e il 2° a retrocedere dal 1888-1889, fissato come origine nella seconda. Quindi dando ad x il valore 4 nella prima e il valore -2 nella seconda, noi avremo il dato teorico del consumo nell'ipotesi che il prezzo del sale fosse rimasto invariato e il dato teorico del consumo a riduzione avvenuta, cioè rispettivamente, a calcoli fatti, quintali 1.579.936 e 1.660.875, la cui differenza è 80.939. Si conclude che la riduzione di 20 lire su 55 non ha accresciuto il consumo che di 80.939 quintali, ossia nella modesta misura di 5,12 per cento.

Questo aumento poi di consumo in parte non fu che trasformazione di un altro consumo, quello del *sale pastorizio*, che non poche famiglie poverissime usavano come surrogato al sale ordinario:

Esercizi finanziari	Consumo di sale pastorizio in quintali	Esercizi finanziari	Consumo di sale pastorizio in quintali	Esercizi finanziari	Consumo di sale pastorizio in quintali
1880-1881	81.476			1886-1887	40.996
1881-1882	85.097			1887-1888	38.838
1882-1883	87.673	1885-1886	60.158	1888-1889	39.390
1883-1884	96.641			1889-1890	38.468
1884-1885	85.130			1890-1891	41.822

Ripetendo il calcolo come sopra, si trova pel 1886-1887, primo anno in cui ebbe pieno effetto la riduzione di prezzo del sale commestibile, una diminuzione di 55.098 quintali nel consumo del sale pastorizio, in confronto del consumo, che si sarebbe probabilmente verificato senza lo stimolo dell'accennata riduzione. Quindi, degli 80.939 quintali di aumento del sale commestibile, 55.098 *rappresenterebbero* una sostituzione della derrata buona ad un cattivo surrogato e i rimanenti 25.841 una vera e propria, per quanto modesta, espansione dell'ordinario consumo.

Abbiamo sottolineato il condizionale appunto per significare che la conclusione è approssimativa soltanto. La interpolazione, invece di fermarsi alla retta, poteva spingersi alla parabola di 2° grado ed anche oltre, con risultati naturalmente diversi da quelli dianzi ottenuti. Tuttavia l'arrestarsi alla retta era giustificato, oltre che dalla brevità delle serie considerate, da un criterio decisivo. Il consumo del sale commestibile pel 1885-1886, esercizio che fu per metà tempo sotto la vecchia tariffa, per l'altra metà sotto la tariffa ridotta, può teoricamente determinarsi prendendo la media aritmetica dei valori che si hanno dalle due equazioni lineari stabilite sopra, quando si faccia nella prima $x = + 3$, nella seconda $x = - 3$. Tale media è di 1.599.349 quintali ed essa si differenzia di soli 2432 quintali, in meno, dal dato osservato (1.601.781). Invece, coll'interpolazione parabolica, la media dei valori teorici pel 1885-1886 sarebbe risultata superiore di 12.233 quintali al dato dell'osservazione. Nessun dubbio quindi che dovevasi preferire l'interpolazione lineare, sebbene anch'essa non conduca alla certezza assoluta del residuo.

La quasi eguaglianza dei coefficienti della x nelle due equazioni è indizio che tutti gli altri antecedenti, presi nel loro insieme, hanno operato in modo press'a poco eguale nei due periodi confrontati e che perciò la loro composizione per numero e specie deve essere rimasta sostanzialmente la stessa.

§ 4. Una combinazione dei tre metodi di eliminazione, dei quali si è parlato finora, può essere esemplificata con una indagine da noi fatta allo scopo di ricavare la *periodicità settimanale* di un fenomeno, di cui son date le variazioni solo *per mesi* (1). Il fenomeno preso in esame è la nuzialità italiana. Orbene gli artifici di eliminazione appaiono tosto applicabili alla soggetta materia, per ciò che i diversi mesi dell'anno non si compongono di un egual numero di giorni dello stesso nome. Se il mese è di 31 giorni e comincia, ad esempio, in sabato, conterrà 5 sabati, 5 domeniche e 5 lunedì, ma solo 4 di ciascun altro giorno; se comincia col martedì, conterrà 5 martedì e altrettanti mercoledì e giovedì, ma solo 4 dei giorni d'altro nome; e così via. Altra circostanza, di cui si deve trarre partito, è che lo stesso mese incomincia in modo diverso per alcuni anni di seguito; ma poi, se la serie d'anni è abbastanza lunga, si ripresenta cogli stessi giorni iniziali e con una periodicità solo perturbata dalla ricorrenza di anni bisestili. Le osservazioni pei 5 mesi sottoindicati, estese al trentennio 1872-1901, dànno il seguente risultato:

---

(1) *Giornale degli Economisti*, ottobre 1904.

MEDIA FREQUENZA DI NOZZE NEI MESI DI 31
(Esclusi il gennaio e il marzo).

Giorno iniziale del mese	Maggio	Luglio	Agosto	Ottobre	Dicembre	Media delle medie
			Medie mensili.			
Lunedì	15.781	11.416	13.242	19.544	19.538	**15.904**
Martedì	16.580	12.376	13.410	19.472	20.209	**16.409**
Mercoledì	15.782	11.684	13.402	18.657	21.181	**16.141**
Giovedì	17.456	12.225	13.321	19.690	21.040	**16.746**
Venerdì	16.531	13.041	14.129	19.729	21.491	**16.984**
Sabato	16.983	12.871	13.531	20.690	22.175	**17.250**
Domenica	17.353	12.729	13.433	19.732	20.362	**16.722**

Raggruppando i primi tre e gli ultimi tre giorni, e prendendo le medie, si ha:

Giorni in cui cominciò il mese	Maggio	Luglio	Agosto	Ottobre	Dicembre
Lunedì, martedì, mercoledì . .	16.048	11.825	13.352	19.224	20.309
Venerdì, sabato, domenica. . .	16.956	12.880	13.698	20.050	21.343
Differenze . . .	— 908	— 1.055	— 346	— 826	— 1.034

Adunque se il mese comincia con uno dei tre primi giorni della settimana, ha una frequenza di nozze sempre inferiore a quella che ha, se incomincia con uno dei tre ultimi. Il gennaio e il marzo, mesi di 31, confermerebbero pure la regola; senonchè essendo essi fortemente perturbati dalla mobilità della Quaresima, la concordanza dei risultati potrebbe parere accidentale; per iscrupolo li abbiamo quindi tenuti fuori conto.

Applicando il metodo di concordanza, si ragionerà così: i 5 mesi in questione presentano tutte le circostanze diverse (temperatura, condizioni sanitarie, genere di produzioni, abitudini del popolo per quanto concerne la scelta dell'epoca delle nozze, ecc.); concordano però in questo che se il mese incomincia con uno dei tre primi giorni della settimana, esso riesce meno frequente di nozze, che non quando incomincia con uno dei tre ultimi. La specie dei giorni iniziali deve dirsi dunque causa o parte necessaria della causa della maggiore o minor frequenza di matrimoni.

Il metodo di differenza l'applicheremo, considerando uno qualunque dei mesi della prima tabella, per esempio, il luglio. Il luglio di un anno somiglia per moltissime circostanze (temperatura, condizioni

sanitarie, genere di occupazioni, ecc.) al luglio di ogni altro anno: anzi, siccome i dati della tabella esprimono medie ciascuno di 4 o 5 termini (1), che devono dar luogo a notevoli compensazioni di variazioni accidentali, così potremo dire *a fortiori* che la media dei mesi di luglio di un gruppo d'anni, somiglia per moltissime circostanze alla media dei mesi di luglio di un altro gruppo d'anni. L'unica circostanza, o almeno la principale, per cui differiscono nel nostro prospetto le medie dei mesi d'uno stesso nome, è la specie del giorno iniziale. Concludesi quindi che il divario di frequenza di nozze per uno stesso mese secondo che esso comincia in un modo o nell'altro, deve attribuirsi precisamente alla qualità del giorno iniziale, ossia alla diversa proporzione dei giorni di un dato nome che esso mese presenta nei diversi casi.

Il metodo dei residui viene ora in acconcio. Tra un mese di 31 giorni cominciante per lunedì e lo stesso mese cominciante per martedì, la differenza è di un lunedì in più e un giovedì in meno, come si può facilmente verificare; similmente tra un mese iniziato in martedì e lo stesso mese iniziantesi in mercoledì, la differenza è di un martedì in più e di un venerdì in meno; e via dicendo. Richiamando dunque i dati dell'ultima colonna del primo prospetto e adottando le lettere l, m, μ, g, ecc. per denotare i singoli giorni della settimana, avremo:

Giorno iniziale	Frequenza media di nozze nel mese	Differenze
Lunedì	15.904	
		$-505 = l - g$
Martedì.	16.409	
		$+268 = m - v$
Mercoledì	16.141	
		$-605 = \mu - s$
Giovedì.	16.746	
		$-238 = g - d$
Venerdì	16.984	
		$-266 = v - l$
Sabato	17.250	
		$+528 = s - m$
Domenica.	16.722	
		$+818 = d - \mu$
Lunedì	15.904	

donde:

$$l = d - 743$$
$$m = d - 741$$
$$\mu = d - 818$$
$$g = d - 238$$
$$v = d - 1009$$
$$s = d - 213$$

La composizione di un mese di 31, cominciante, ad esempio, per lunedì, si potrà esprimere così:

$$5l + 5m + 5\mu + 4g + 4v + 4s + 4d = 15{,}904.$$

E sostituendo ad l, m, μ, ecc. i loro valori per rispetto a d, si ottiene per d il numero 1,073, e quindi la media frequenza dei diversi giorni della settimana risulta per i 5 mesi considerati la seguente:

---

(1) Infatti, per prendere un esempio, nel trentennio 1872-1901 il luglio cominciò cinque volte per lunedì, cinque per domenica e quattro per ciascuno degli altri giorni.

Giorni	Media frequenza di nozze
Lunedì	320
Martedì	332
Mercoledì	255
Giovedì	835
Venerdì	64
Sabato	860
Domenica	1073

Lo stesso calcolo può ripetersi, *mutatis mutandis*, per i mesi di 30 giorni e con particolari avvedimenti si possono combinare le due serie ottenute. Ne esce confermato in cifre il fatto, vagamente noto per osservazione comune, di una agglomerazione di nozze nel giovedì, nel sabato e soprattutto nella domenica e di una scarsità eccezionale nel venerdì, imputabile al pregiudizio popolare, che considera questo giorno come nefasto.

Il lettore naturalmente non vorrà guardare all'irrilevanza della ricerca in sè stessa, ma al problema metodologico, che essa involge e che non si risolve se non con una opportuna combinazione dei metodi di *concordanza*, di *differenza* e dei *residui*.

§ 5. Da ultimo, quando non è possibile uno dei detti processi di eliminazione, si ricorre al metodo delle variazioni concomitanti, che così si enuncia:

Se un fenomeno varia costantemente in una determinata direzione e misura, allorchè un secondo varia in altra determinata direzione e misura, questo si dirà causa di quello, oppure entrambi si diranno dipendenti da una causa comune, che li fa variare in modo armonico.

Questo metodo ha per base il principio che ogni modificazione della causa è seguita da una modificazione dell'effetto — principio che non è esso medesimo se non una generalizzazione induttiva.

L'influenza della luna sulle maree non potrebbe essere accertata coi metodi di eliminazione, perchè non è in nostra facoltà di predisporre uno sperimento, in cui si abbia assoluta assenza dell'astro perturbatore; però osservando che ad ogni variare di posizione della luna corrisponde costantemente il variare di posizione dell'alta marea, concludiamo che i due fatti sono collegati tra loro da un rapporto causale.

Noi non possiamo eliminare del tutto gli attriti e resistenze, che producono l'arresto di un pendolo messo in oscillazione; però osservando che man mano si diminuisce artificialmente l'attrito sull'asse di sospensione, la resistenza dell'aria (col rarefarla), ecc., si prolunga il tempo nel quale il pendolo rimane in moto, concludiamo che le dette circostanze sono la vera causa dell'arresto del pendolo e, per via deduttiva, che soppresse totalmente quelle, il pendolo seguiterebbe ad oscillare senza arrestarsi mai.

Il metodo delle variazioni concomitanti ha grande importanza nello studio dei fenomeni collettivi. Il Mill lo contesta. Se le cause operanti

in una società, egli dice, producessero effetti differenti una dall'altra nella specie, se la ricchezza dipendesse da una causa, la pace da un'altra, la moralità pubblica da una terza, ecc., distintamente, noi potremmo, pur nell'impossibilità di isolare sperimentalmente le cause l'una dall'altra, attribuire a ciascuna il suo effetto, osservando le variazioni concomitanti di questo e di quella. Ma ogni aspetto della vita sociale è influenzato da innumerevoli cause; e la mutua azione degli elementi coesistenti della società è tale che qualunque cosa alteri l'uno di essi, se non modifica gli altri direttamente, li modifica indirettamente..... Si risponde che, appunto perchè i fatti sociali sono estremamente complessi e variabili, solo che essi siano suscettivi di osservazione statistica daranno luogo, graficamente rappresentati, alle curve più disparate per il numero, per la grandezza e per l'ordine delle loro variazioni. Perciò se ci accade di trovare due curve molto concordanti nelle particolarità del loro andamento, sarà assai improbabile che la concordanza sia accidentale, assai probabile invece che dipenda da un rapporto di causa ad effetto, che lega insieme i fenomeni considerati.

Concomitanza non vuol dire *sincronismo* di variazioni; tra la causa e l'effetto può correre un intervallo più o meno lungo. Ad un anno di nascite molto scarse corrisponderà solo in capo ad un ventennio un anno di scarso numero d'inscritti alla leva; ad un periodo di crescente frequenza dei fanciulli alle scuole elementari corrisponderà solo dopo tre o quattro lustri una forte proporzione d'iscrizioni nelle liste elettorali o una forte proporzione di sposi capaci di firmare l'atto nuziale.

Concomitanza non vuol dire nemmeno *egual direzione delle variazioni*. Il nesso tra i fenomeni può essere tale che, crescendo l'uno, abbia a diminuire l'altro, e viceversa. La natimortalità cresce col crescere dell'età delle madri, ma la fecondità invece diminuisce. La proporzione dei beni mobili cresce col crescere dell'entità dei patrimoni; la proporzione degli immobili invece diminuisce. L'importazione di monete o di metalli preziosi monetabili tende a crescere col diminuire del corso dei cambi, l'esportazione invece diminuisce. E così via.

Ancora: la concomitanza delle variazioni non implica che queste debbano essere *di eguale grandezza*. L'un fenomeno può variare nella ragion dei quadrati, dei cubi, delle radici quadrate, ecc., o in un'altra ragion determinata, mentre l'altro varia in ragione semplicemente aritmetica. Così stando alla vecchia legge empirica di Gregorio King, il rincaro del grano seguirebbe in questa misura alle deficienze dei raccolti:

Deficienza in confronto del raccolto medio	Rialzo del prezzo preso come unità il prezzo normale
1 decimo	3 decimi
2 decimi	8 »
3 »	16 »
4 »	28 »
5 »	45 »

nel qual caso alle variazioni in ragione aritmetica del primo fenomeno corrisponderebbero variazioni inverse del secondo, procedenti secondo una curva di terzo grado.

Può darsi che due (o più) fenomeni siano in rapporto causale con un terzo. Allora quello dei due, le cui variazioni concordano meglio colle variazioni del terzo, deve ritenersi legato da questo da un nesso più stretto e diretto.

Esempio: le legittimazioni di figli naturali per susseguente matrimonio dei genitori variano più concordemente col variare delle nascite di illegittimi riconosciuti dai genitori, che non col variare delle nascite di illegittimi esposti o, comunque, non riconosciuti. Se ne inferisce che le legittimazioni ridonano una famiglia legale assai più spesso ai primi che ai secondi; in altre parole, gli stessi motivi, che già determinarono i genitori naturali al riconoscimento del figlio al momento della denunzia della nascita allo stato civile, li predispongono poi in favore della legittimazione per susseguente matrimonio.

Ma si dà pure il caso che le variazioni di un fenomeno siano determinate dal concorso delle variazioni di due o più altri fenomeni. Allora la concomitanza delle serie, distintamente prese, può mancare e il metodo in esame è in difetto. Le cause, invece che risultare dal confronto delle serie, devono essere presunte o ammesse in base ad altre forme di osservazione o di ragionamento. Una volta ammesse, con un calcolo di correlazione doppia o tripla, noi miriamo ad attribuire a ciascuna la parte, che più probabilmente le spetta nell'effetto.

Qui torna opportuna una considerazione, che abbiamo passata sotto silenzio nelle pagine dedicate alla *Teoria delle correlazioni*. La considerazione è questa. Allorchè il fenomeno A si presume obbedire all'influenza combinata di B e di C, noi determiniamo anzitutto la correlazione di A con B e la correlazione di C con B. Ciò fatto, paragoniamo tra loro le oscillazioni residue, che tanto la serie A, come la serie C presentano, non spiegabili coll'influenza di B. È come dire, insomma, che le variazioni di C non si fan pesare in calcolo per quella parte di lor medesime, che è in perfetta concomitanza colla serie B. La qual cosa è giustificabile, quando sia dimostrato che tra B e C intercede pure un rapporto causale. Ma se B e C fossero l'un dall'altro indipendenti, se la concomitanza più o meno completa delle loro variazioni fosse accidentale, l'eliminarla nel calcolo di correlazione per A sarebbe arbitrario. La via logica in tal caso è quella di calcolare distintamente la correlazione tra A e B e tra A e C e prendere la media delle equazioni così ottenute. Sia:

$$A = K + m \,.\, \delta B$$

$$A = K + n \,.\, \delta C.$$

Si avrebbe:

$$A = K + \frac{m \,.\, \delta B + n \,.\, \delta C}{2}.$$

Meglio ancora, in luogo della media aritmetica semplice, fare la media ponderata, attribuendo a ciascuna equazione un coefficiente di importanza proporzionato alla precisione dei risultati, che essa dà; in altri termini, inversamente proporzionale allo scostamento medio tra i dati dell'osservazione e i dati teorici forniti dalle equazioni.

§ 6. Da questa rapida rassegna dei metodi di induzione sperimentale emerge chiaro che nell'ordine dei fatti collettivi essi non possono dar luogo se non a conclusioni probabili. La probabilità può essere accresciuta dalla conferma dei risultati ottenuta coll'applicazione successiva o combinata dei diversi metodi. Nell'enumerazione delle circostanze comuni o differenziali tra due serie, se non possiamo arrivare alla convinzione che una data circostanza è l'*unica* comune o l'unica diversa, basterà la convinzione ragionata che tale o tal'altra circostanza è la *principale* per cui si assomigliano o si differenziano le serie. Noi non dobbiamo escludere delle conoscenze approssimative, per ciò solo che sono approssimative; la loro introduzione come elementi di calcolo nella soluzione di speciali problemi, si chiarisce in molti casi utilissima.

L'enumerazione delle circostanze, che possono essere antecedenti necessari o non necessari di un fenomeno, richiede dallo statistico cognizioni di vario ordine, spirito critico e attività d'immaginazione. Diciamo attività d'immaginazione, perchè spesso si deve partire dalla *supposizione* di una circostanza, per concludere poi in base ai fatti per la sua ammissibilità o inammissibilità. Naturalmente non è possibile teorizzare in materia, nella quale il successo dipende dalle qualità personali dello studioso. Solo ci piace chiudere questo capitolo con uno schema di enumerazione, ridotto per ragioni di spazio ai minimi termini, prendendo ad esempio le variazioni nel rapporto di mercato tra l'oro e l'argento durante il quarantennio 1862-1902.

ANNI	Per 1 Kgr. d'oro quanti d'argento	Circostanze principali enumerate
1862-1865	15,40	Lega monetaria latina — Produzione rallentata dell'oro, aumentata per l'argento.
1866-1870	15,55	Corso forzoso in Italia e, alla fine del periodo, in Francia — Produzione in lieve aumento per l'oro, in aumento più sensibile per l'argento.
1871-1873	15,58-15,93	Adozione del tipo oro in Germania, Stati scandinavi e Stati Uniti del Nord America — Produzione rallentata dell'oro, in forte aumento per l'argento.
1874-1875	16,16-16,63	Limitazione delle coniazioni d'argento per parte della Lega latina — Adozione del tipo oro in Olanda — Vendite di argento smonetato in Germania.

(*Segue*) Schema d'enumerazione.

ANNI	Per 1 Kgr. d'oro quanti d'argento	Circostanze principali enumerate
1876-1880	17,80-18,06	Coniazioni d'argento sospese dalla Lega latina — Abolizione del corso forzoso in Francia — Il *Bland-bill* agli Stati Uniti — Regresso nella produzione dell'oro; progresso in quella dell'argento.
1881-1885	18,24-19,39	Abolizione del corso forzoso in Italia — Rinnovazione della Lega latina — Ancora regresso nella produzione dell'oro e progresso in quella dell'argento.
1886-1889	20,78-22,09	Rapido progresso della produzione dell'argento — Inchiesta monetaria inglese.
1890	19,77	Legge Windom, che aumenta gli acquisti obbligatorii d'argento per parte del Tesoro, negli Stati Uniti.
1891-1892	20,92-23,67	Riforma monetaria in Austria-Ungheria — Conferenza monetaria di Bruxelles — Aumento nella produzione d'entrambi i metalli.
1893	26,47	Sospensione delle coniazioni d'argento nell'India inglese — Abrogazione della legge sopra gli acquisti di argento da parte del Tesoro agli Stati Uniti.
1894-1896	32,59-30,67	Tassa del 5 % sull'argento importato in India — Crisi monetaria agli Stati Uniti — Crisi dei valori minerarii al Transvaal — Adozione del tipo oro nel Chilì.
1897-1902	35,50-39,19	Adozione del tipo oro in Russia, Giappone e India — Forte aumento nella produzione d'ambo i metalli — Guerra anglo-boera.

CAPO TERZO

L'ufficio delle ipotesi nella Statistica.

Sommario: § 1. Le ipotesi e la Statistica congetturale. — § 2. Ammontare del debito ipotecario in Italia. — § 3. La mortalità per febbre puerperale e la fecondità delle coniugate alle varie età. — § 4. Congetture del Beloch sulla popolazione di Roma antica. — § 5. La classificazione dei libretti di risparmio postali. — § 6. Relazioni tra reddito e patrimonio. — § 7. Le particolarità demografiche delle provincie ex-pontificie. — § 8. Conclusione.

§ 1. L'ipotesi è l'enunciazione di un principio, ammesso il quale e combinato con premesse minori stabilite per osservazione, si arriva a certe conseguenze. Se le conseguenze, immediate o mediate, sono in accordo coi fatti, l'ipotesi si dice *reale; irreale* nel caso contrario.

Il carattere del ragionamento fondato su ipotesi non è sempre necessariamente deduttivo; è tale solo quando il principio ammesso trascende la possibilità dell'osservazione. I concetti del punto inesteso in geometria, del corpo perfettamente rigido in meccanica, dell'infinitesimo nel calcolo, dell'etere in fisica, ecc., sono veramente ipotesi di tal genere e caratterizzano il ragionamento deduttivo. Quando invece il principio ammesso è esso medesimo una generalizzazione induttiva, una estensione della regola di casi osservati a casi non osservati, ma osservabili, e noi l'impieghiamo per evitare le difficoltà di ulteriori ed esaurienti osservazioni, il ragionamento, che il Mill direbbe *deduttivo-concreto*, non è in ultima analisi se non una variante dell'ordinaria induzione.

Nel corso di quest'opera ci venne fatto più volte di segnalare l'uso di ipotesi nella soluzione di problemi statistici. Così l'Halley si è valso dell'ipotesi di una popolazione stazionaria per costruire una tavola di sopravvivenza colla sola scorta di una classificazione dei morti per età. Il Livi ha fatto l'ipotesi del miscuglio di due gruppi etnici differenti sino a certi limiti per statura media, per dimostrare che entro tali limiti la curva del miscuglio non può presentare due vertici. Nei problemi relativi alla scelta matrimoniale noi abbiamo confrontato i dati della osservazione con quelli teorici desunti dall'ipotesi di combinazioni come in un giuoco di sorte. Nei procedimenti di perequazione e d'interpolazione si applica largamente il principio — in gran parte ipotetico — che le funzioni discontinue presso l'individuo o presso i gruppi scelti tendono a farsi continue nella massa, quando questa risulta da una sufficiente varietà di soggetti e di situazioni. Così dicasi del principio della compensazione degli errori accidentali nelle grandi masse di osservazioni, ecc.

Tutta la STATISTICA CONGETTURALE si basa sovra enunciati, che hanno ragion d'essere in criterii di *analogia*, di *proporzionalità*, di *continuità*, di *concomitanza*, ecc., dei fenomeni, e la cui condizione generale è che involgano un *minimum* di arbitrio. Anche nella più completa assenza di osservazioni dirette riguardo a un dato fenomeno, non si può dir completa l'ignoranza nostra in materia, se noi sappiamo almeno quel che succede di altri fenomeni connessi in qualche modo col primo e se possiamo stabilire, anche per approssimazione, la natura del legame che intercede fra l'uno e gli altri. Qui molto dipende dalle particolari attitudini dello studioso; una teoria generale non è facile a concepirsi. Pertanto il procedere per via d'esempi, se appare difettoso dal punto di vista sistematico, ha in compenso un valor suggestivo, che nell'interesse didattico non bisogna trascurare.

§ **2.** Sia da determinare il debito ipotecario fruttifero che grava sulla proprietà di private persone in Italia. La situazione di questo debito, senza distinzione delle persone dei debitori, fu accertata con metodo diretto una volta sola, nel 1871, e dopo d'allora si è calcolata anno per anno mediante la semplice addizione delle ipoteche nuovamente inscritte e la sottrazione delle ipoteche ridotte o cancellate. Senonchè è noto come, massime per i piccoli debiti, la cancellazione dell'ipoteca non segua immediatamente all'estinzione del debito, ma si ritardi, a cagione della spesa e del disagio che importa, finchè non sorga l'interesse a dimostrare la libertà dell'immobile o finchè per il decorrere del tempo l'ipoteca non venga cancellata d'ufficio per perenzione. D'altronde gli stessi conservatori delle ipoteche non sono sempre in grado di riconoscere l'attinenza che più iscrizioni possono avere con un solo e medesimo debito; donde la probabilità di duplicazioni di somme. La situazione calcolata nel modo che si è detto sembrava quindi aggravarsi d'anno in anno per un doppio errore: per l'aggiunta di somme in più del vero e per la sottrazione di somme inferiori al vero. Al 31 dicembre 1896 il carico ipotecario apparente — senza distinzione delle persone fisiche o morali al cui nome era inscritto — oltrepassò così i 10 miliardi (1); 4 miliardi più che nel 1871.

Le statistiche delle successioni permettono ora di risolvere il quesito proposto, pur col sussidio di qualche ipotesi. Esse denunziano una media di 85 milioni di debiti ipotecarii gravanti gli immobili trasmessi ogni anno *mortis causa*; il qual dato concorda con quello delle rendite e crediti garantiti da ipoteca, che figurano annualmente tra le attività ereditarie. Gli 85 milioni in parola arrivano a circa 100, se si tien

(1) Oggi questa cifra tende nelle stesse statistiche del debito ipotecario a diminuire con rapidità, perchè trascorso il trentennio dall'applicazione del codice civile, le ipoteche non cancellate per vecchi debiti estinti, vengono ora man mano a cadere in perenzione e sono cancellate d'ufficio.

conto in primo luogo delle passività assorbenti l'attivo delle successioni oberate, e in secondo luogo degli oneri accollati ai donatarii. Quale percentuale rappresentino i 100 milioni di debito ipotecario ragguagliati al valore degli immobili trasmessi, è facile determinare, riflettendo che fra successioni (comprese le oberate) e donazioni (considerate come anticipi di eredità) si ha una massa annua di 800 milioni di lire per terreni e fabbricati, elevabile a 900 per tener conto, secondo le stime dei pratici, delle somme che sfuggono al fisco per dichiarazioni di valore inferiori al vero. Il debito ipotecario rappresenta dunque non più dell'11 % del valore lordo degli immobili posseduti da privati individui, autori di successioni e donazioni (1).

Rimane a vedere qual è in assoluto l'ammontare di esso debito sulla proprietà immobiliare dei viventi. La risposta che si dà a questa domanda è la stessa che si dà pel calcolo della ricchezza privata in genere: l'intervallo medio tra due generazioni successive, ossia il tempo in capo al quale i beni oggi ricevuti saranno di nuovo trasmessi agli eredi degli attuali eredi, può stimarsi di 36 anni; moltiplicando dunque il medio ammontare annuo delle successioni (e donazioni) per 36, si avrà il totale della ricchezza esistente in mano di privati individui.

Questo ragionamento è da accogliersi, quantunque trascuri una circostanza di qualche rilievo. Esso tratta gli eredi attuali, come se fossero assolutamente sprovvisti di patrimonio proprio, da loro medesimi creato all'infuori di quello che ricevono o già possono aver ricevuto per via ereditaria. Mentre è certo che accanto alla ricchezza ereditata può darsi una ricchezza non ereditata, dovuta all'opera personale di coloro che ora rivestono anche la qualità di eredi. Se questi fossero in maggioranza dei minorenni, che non ebbero ancora tempo e modo di formarsi un patrimonio indipendentemente dalla congiuntura di una eredità o di una donazione, il metodo sopra indicato sarebbe corretto ed esauriente; ma siccome si può divenire eredi o donatarî in qualsiasi epoca della vita e l'età media, in cui lo si diventa, batte intorno ai 34 anni (2), è chiaro che molti, prima ancora di rivestire la qualità di eredi, possono aver bene impiegati gli anni dalla maggiore età ed essersi formato *ex novo* un patrimonio proprio. Poca o molta che sia questa ricchezza d'origine non ereditaria, che ogni generazione aggiunge alla

(1) Questo calcolo fu da noi pubblicato già nella *Zeitschrift f. Socialwissenschaft*, di J. Wolf, anno II, fasc. 4°, Berlino 1899. La nostra Direzione generale del Demanio (*Boll. di Statistica e Legislazione comparata*, anno III, fasc. 1°) ha seguìto sostanzialmente lo stesso metodo.

(2) Gli eredi sono assai di rado dei minorenni. Se lo fossero di frequente, data la loro notevole mortalità, essi dovrebbero figurare spesso come autori di successioni e dovrebbero pure trovarsi inscritti in buon numero nei ruoli dei contribuenti alle imposte dirette. La qual cosa non si verifica. Nello scritto citato nella nota precedente dimostrai che su 100 successioni, 4 o 5 soltanto hanno per autori individui minorenni; e quanto alle iscrizioni nei ruoli dei

ricchezza ricevuta dai padri, essa dovrà pure, moltiplicata per 36, integrare il computo di cui sopra si è parlato. Ad ogni modo resta inteso che se si istituisce il calcolo della ricchezza privata in base ai soli dati delle trasmissioni ereditarie (donazioni comprese), bisognerà considerare il moltiplicatore 36 piuttosto come un buon minimo, che come una buona media (1).

Ciò premesso, la ricchezza immobiliare trasmessa ogni anno *mortis causa* o a titolo gratuito *inter vivos* nella somma rotonda di 900 milioni, costituirebbe $^1/_{36}$ dell'intera ricchezza immobiliare posseduta da privati individui, la quale pertanto ammonterebbe ad almeno 32 miliardi e mezzo. E l'onere ipotecario gravante su di essa nella ragione dell'11 % salirebbe ad almeno 3 miliardi e 560 milioni di lire. In difetto di migliori elementi di stima, questa valutazione può ritenersi abbastanza approssimata.

§ **3.** Le statistiche demografiche italiane non fanno conoscere per tutto il regno come varii coll'età la fecondità delle donne maritate. È possibile con qualche ipotesi supplire alla lacuna delle osservazioni? Non dico delle ipotesi *per analogia*, che si potrebbero comodamente avanzare, proponendo, ad es., di estendere a tutta Italia i coefficienti accertati per alcuni comuni urbani (come Udine, Roma, Cremona) e

---

contribuenti, quelle di minorenni non arrivano all'1 $^1/_2$ per cento della popolazione sotto i 21 anni. La grande maggioranza degli eredi dev'essere dunque di età maggiore.

L'età normale dei morti, determinata col noto metodo del Lexis (V. *Principii di Demografia*, pag. 45 e segg.), è in Italia di 74 anni. Poco inferiore deve essere l'età media degli autori di successioni; mettiamo, 70 anni. Togliendo da questo numero gli anni di durata di una generazione, si ha per residuo 34, età media degli eredi.

(1) Che sia un buon minimo, anzi che una buona media, si argomenta pure da un calcolo basato stavolta su una ipotesi di *proporzionalità*. La ricchezza privata è quasi tutta nelle mani della popolazione maggiorenne, potremmo dire della popolazione da 25 anni in su. Convien partire dal limite di 25 anni, piuttosto che da quello di 21, perchè appunto intorno ai 25 i figli di famiglia cominciano ad essere proprietari in proprio nome, sia che ricevano dai parenti assegni per impianto di negozio, doti per matrimonio, ecc., sia che loro succedano come eredi. Questo secondo caso deve verificarsi con una certa frequenza per figli venticinquenni, i cui genitori, in media sessantenni, vanno incontro alle alte quote di mortalità. Ora i morti da 25 anni in su sono la 43ª parte dei viventi, pure da 25 anni in su. Se i primi possedevano, prima di passare a miglior vita, 900 milioni di immobili, i secondi, ammesso il criterio della proporzionalità, possederanno 43 volte 900 milioni ossia quasi 39 miliardi. Il moltiplicatore sarebbe dunque 43 e non 36. Senonchè la proporzionalità non può essere rigorosa e il moltiplicatore 43 pecca per eccesso, come l'altro pecca per difetto. Infatti l'osservazione comune apprende che la ricchezza posseduta cresce col crescere dell'età, almeno fino ad una certa epoca della vita; ora la proporzione dei vecchi tra i morti è maggiore che non tra i viventi. Io credo quindi che ci avvicineremmo al vero adottando un moltiplicatore intermedio fra 43 e 36, per esempio, il 40.

aggiustandoli in modo da far corrispondere il risultato al totale annuo delle nascite legittime dell'intero paese. No; intendesi parlare qui di ipotesi che riguardino *fatti concomitanti* alla fecondità, osservati nelle diverse categorie d'età delle donne coniugate.

Tra i fatti concomitanti alla fecondità c'è la *mortalità per febbre puerperale* e per altre malattie di parto e puerperio. Questo fenomeno è illustrato nelle *Statistiche delle cause di morte* colla desiderata distinzione delle donne per età. L'ipotesi che tosto si affaccia è questa: la febbre puerperale, avendo carattere di *causa accidentale*, colpisce probabilmente le puerpere in proporzioni uguali, qualunque sia l'età; in altri termini, devesi ritenere che questa malattia infettiva non colpisca di preferenza le donne giovani e risparmii le mature, o viceversa. Lo stesso non potremmo dire di altre malattie inerenti allo stato di fecondità. Ad es.: la metrorragia dopo il parto dev'essere, a parità d'altre circostanze, più facile nelle puerpere non giovani, perchè coll'età diminuisce la resistenza dei tessuti dei vasi sanguigni. Limitiamoci dunque alla febbre puerperale. Il seguente prospetto ci dà anzitutto la classificazione delle coniugate per età al 10 febbraio 1901 (1); indi la classificazione per età delle donne morte per febbre puerperale, in media annua, nel quadriennio 1899-1902, le quali donne sono per la massima parte coniugate (2); infine il numero medio annuo delle nascite legittime per lo stesso periodo, ma senza la distinzione per età delle madri, che appunto si tratta di stabilire in base alla accennata ipotesi:

ETÀ	Coniugate esistenti al 10 febbraio 1901	Donne morte per febbre puerperale (media annua)	Nascite legittime (nati vivi e nati morti insieme)
15-20	95.618	38	?
20-25	541.036	246	?
25-30	777.656	262	?
30-35	822.388	241	?
35-40	786.979	186	?
40-45	726.831	103	?
45-50	620.885	14,75	?
	4.371.393	1.090,75	1.056.107

(1) Il censimento del 1901 porta la classificazione per anni di nascita, corrispondenti ai gruppi d'età da 15 a 18, da 18 a 21 e da 21 a 25 con un eccesso di 40 giorni, dovuto al fatto che il censimento non ebbe luogo al 1° gennaio, ma al 10 febbraio. Mediante una interpolazione parabolica di 2° grado abbiamo ricostituita la classificazione per gruppi quinquennali di età, ma senza tener conto dell'eccesso notato di 40 giorni.

(2) Le nubili e le vedove rappresentano, secondo i dati del 1899 e 1900, meno del 9 % delle donne morte per febbre puerperale.

Porremo la questione così: se la febbre puerperale colpisce e miete in eguali proporzioni, cioè senza riguardo all'età, le donne in istato di puerperio, e se per 1.056.107 nascite legittime si ebbero circa 1091 donne morte per tal causa in età da 15 a 50 anni, per quante nascite legittime si saranno verificate le 38 morti di donne in età da 15 a 20? per quante altre nascite legittime si saranno verificate le 246 morti di donne in età da 20 a 25 anni? E così via. La regola del tre dà per l'incognita i valori 36.793, 238.187, ecc., come si vede dal prospetto seguente:

ETÀ	Coniugate esistenti	Numero probabile dei nati	Nati per 100 coniugate
15-20	95.618	**36.793**	**38,5**
20-25	541.036	**238.187**	**44,0**
25-30	777.656	**253.679**	**32,6**
30-35	822.388	**233.346**	**28,4**
35-40	786.979	**180.093**	**22,9**
40-45	726.831	**99.728**	**13,7**
45-50	620.885	**14.281**	**2,3**
	4.371.393	1.056.107	24,2

I coefficienti trovati si accordano così bene con ciò che sappiamo per osservazione comune e per indagini parziali di alcuni nostri centri urbani o generali d'altri paesi, che l'ipotesi proposta dev'essere considerata come attendibile e reale (1). Irreale sarebbe risultata, se l'avessimo estesa alle altre malattie inerenti allo stato di fecondità, come si può vedere da un calcolo condotto con gli stessi criteri del precedente:

ETÀ	Coniugate esistenti	Donne morte per altre malattie di gravidanza, parto e puerperio	Numero dei nati calcolato come nel prospetto precedente	Nati per 100 coniugate
15-20 . . .	95.618	50,50	**27.384**	**28,6**
20-25 . . .	541.036	302,25	**163.717**	**30,3**
25-30 . . .	777.656	374,50	**202.853**	**26,1**
30-35 . . .	822.388	425,75	**230.613**	**28,0**
35-40 . . .	786.979	475,50	**257.561**	**32,7**
40-45 . . .	726.831	269	**145.707**	**20,0**
45-50 . . .	620.885	52,25	**28.502**	**4,6**
	4.371.393	1.949,75	1.056.107	24,2

(1) I coefficienti di natalità risulterebbero alquanto più elevati per le coniu-

Qui la serie dei coefficienti, in contrasto con tutto ciò che sappiamo per altre vie, darebbe una fecondità piuttosto costante fino al 40° anno e decisamente decrescente solo nell'intervallo fra i 40 e i 50 anni d'età. L'ipotesi proposta non potrebbe dunque estendersi all'insieme delle altre malattie inerenti allo stato di fecondità; ma deve limitarsi alla febbre puerperale (1).

§ 4. Nelle sue indagini sulla *Popolazione di Roma antica*, Julius Beloch ci ha dato un vero modello di statistica congetturale in materia storica (2). Gli elementi di calcolo sono vari e non possono dissociarsi da ipotesi fondate or sul criterio di analogia, or su quello di proporzionalità, ecc. Si conosce dal testamento politico di Augusto il numero dei cittadini; si conosce l'estensione della città e approssimativamente quella delle sue singole regioni; è noto il numero delle case all'epoca costantiniana e si ha qualche informazione sul consumo di grano. Incognito invece il numero degli schiavi, che il Beloch suppone nel rapporto di 1 per 3 abitanti (1 schiavo ogni 2 liberi), come si aveva a Pergamo. La scarsezza delle iscrizioni mortuarie non latine è dall'autore presa come indizio della scarsezza dei peregrini. Quanto alla proporzione dei sessi nella popolazione, l'elemento maschile immigrato doveva essere in notevole preponderanza, attratto dalle distribuzioni gratuite di grano, da cui erano invece escluse le femmine. Combinando questi ed altri elementi di calcolo con prudenti ipotesi, il Beloch determina la popolazione di Roma imperiale nei limiti di 945.000 a 1.035.000 abitanti, correggendo le valutazioni fantastiche di taluni scrittori.

---

gate giovani e alquanto più ridotti per le altre, se tenessimo conto della circostanza che i gruppi di età, desunti da quelli degli anni di nascita, sono a dir vero in eccesso di 40 giorni. Così il gruppo indicato da 15 a 20 anni va realmente da 15 anni e 40 giorni a 20 anni e 40 giorni; lo stesso dicasi degli altri gruppi. La correzione, che all'uopo si dovrebbe introdurre nella classificazione importerebbe una lieve diminuzione dei primi tre gruppi e un lieve aumento degli ultimi quattro; i quozienti di fecondità invece ne verrebbero modificati in senso contrario. Ma il piccolo errore, che si commette, trascurando l'eccesso dei 40 giorni è compensato dal fatto che le donne morte per febbre puerperale non sono tutte coniugate o vedove, ma in piccola parte nubili, tra le quali sono relativamente numerose le giovani, più che nol siano tra le coniugate.

(1) Per il triennio 1887-1889 la *Statistica delle cause di morte* ha distinto il gruppo delle altre malattie di gravidanza, parto e puerperio; dai dati d'allora parrebbe che la pelvi e metroperitonite puerperale seguano lo stesso andamento della febbre puerperale; invece la metrorragia e la sincope dopo il parto, l'eclampsia e la distocia colpiscono le puerpere d'età matura in proporzioni più forti che non colpiscano le giovani.

(2) V. *Bulletin de l'Institut international de Statistique*, Tome I, 1re et 2me Livraison, 1886. La citata ricerca con altre consimili sulla popolazione del mondo greco-romano si trova pure nell'opera: *Die Bevölkerung der griechisch-römischen Welt*, Leipzig, Duncker und Humbolt, 1886, dello stesso autore.

§ 5. Anche dell' ipotesi della *continuità* si fa o si può fare largo uso nella statistica congetturale. Quando supponiamo che la popolazione di un paese a partire da un censimento si accresca anno per anno di una quantità eguale a quella accertata nell' intervallo tra il censimento in questione ed uno anteriore, noi utilizziamo l' ipotesi della continuità degli incrementi. Non altrimenti, quando inscriviamo nel preventivo di un'entrata una maggior somma pari all' incremento medio risultante dagli ultimi consuntivi. In breve, nell' ignoranza in cui possiamo essere circa i mutamenti intervenuti nel concerto delle cause, da cui dipende un fenomeno, ci atteniamo alla norma logica del minimo arbitrio col supporre invariato il concerto o sistema delle cause.

L' ipotesi della continuità è, come vedemmo, a base dei procedimenti interpolatorii. Conosciuto un tratto della curva, il resto della figura è completamente, in linea teorica, determinato. Se, dunque, gli elementi di una seriazione statistica omogenea fossero conosciuti tutti per un certo tratto della scala di misura e conosciuti solo in parte pel tratto rimanente, prolungando il primo tratto di curva potremmo integrare la seriazione anche là dove essa, per difetto di osservazioni statistiche, presenta una lacuna. Il caso che qui illustriamo varrà meglio di ogni enunciato generico.

Al 31 dicembre 1893 i *libretti di deposito a risparmio* presso le nostre casse di risparmio ordinarie, società ordinarie di credito e banche popolari, erano in numero di 1.894.306 per una somma di lire 1537 milioni. Di tali libretti si conosceva anche la classificazione per valore:

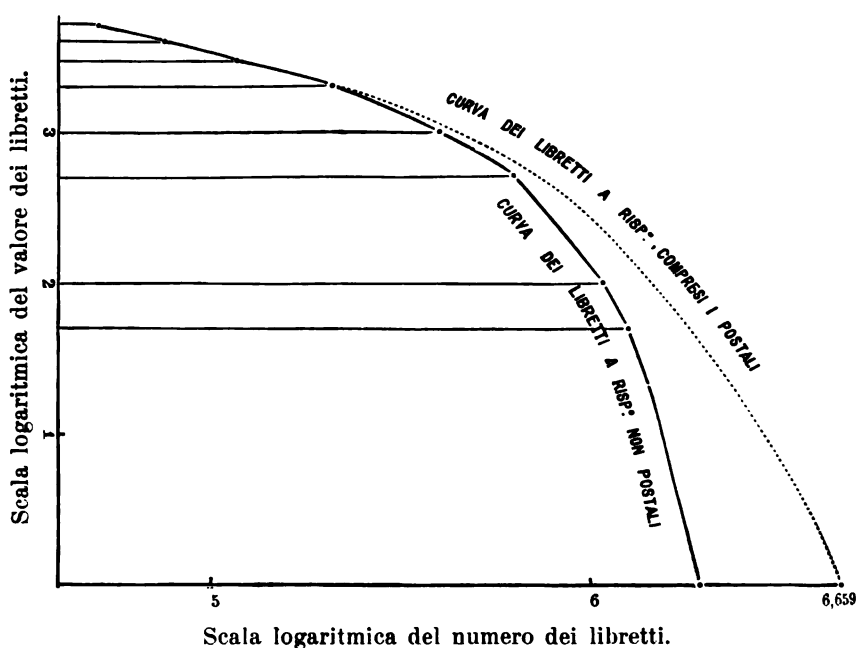
CLASSI DI VALORE	Numero dei libretti a risparmio presso tutti gli Istituti escluse le Casse postali	Somma media per libretto (lire)
Meno di 50 lire	611.042	20,26
Da 50 a 100 lire	205.024	73,45
» 100 a 500 »	451.531	278,10
» 500 a 1000 »	218.968	724,27
» 1000 a 2000 »	199.485	1422,57
» 2000 a 3000 »	91.132	2381,86
» 3000 a 4000 »	42.303	3402,27
» 4000 a 5000 »	25.703	4399,05
Oltre 5000 lire	49.118	9514,43
Totale . . .	1.894.306	Media gen. 811,26

Al 31 dicembre 1893 anche le casse postali di risparmio noveravano ben 2.673.127 libretti per una somma di 400 milioni (media per

libretto = L. 149.65); dei quali libretti però non è data la ripartizione per valore. Solo si sa che essi non discendono, salvo per eccezione, sotto il limite di 1 lira e non salgono, salvo per eccezione, oltre il limite di 2000 lire, perchè da questo punto in avanti è stabilito che l'eccedenza del deposito non frutti interesse. Aggiungasi che nel corso di un medesimo anno non si possono varcare, con diritto a interesse, le 1000 lire di deposito, la qual disposizione deve rendere poco numerosa la classe di libretti fra 1000 e 2000 lire. Il problema che ci proponiamo è il seguente: si può, colla scorta di qualche ragionevole ipotesi, determinare la probabile classificazione per valore anche dei libretti postali a quella data?

L'ipotesi di *proporzionalità* si chiarirebbe subito irreale. Se cioè noi supponessimo ripartiti i libretti postali nelle stesse proporzioni in cui lo sono quelli delle altre casse nei limiti da 1 a 2000 lire, noi otterremmo, applicando convenienti valori medii per ogni categoria, una somma totale di depositi assai superiore ai 400 milioni ricordati. Tale ipotesi va dunque scartata.

Proviamone un'altra, l'ipotesi della *continuità*, e traduciamola ad effetto in un diagramma a doppia scala logaritmica.

Scala logaritmica del numero dei libretti.

Sommando il numero dei libretti postali con quelli delle altre casse si ha un totale di 4.567.433, al qual numero corrisponde il logaritmo 6,65967. Fissato questo punto sull'asse delle ascisse, congiungiamolo con una curva tracciata ad occhio coll'estremo della ascissa, che dà il

logaritmo del numero di libretti superiori a 2000 lire. Tale curva (punteggiata nel diagramma fino al limite di 2000 lire, e indi continua) dovrebbe rappresentarci la seriazione completa dei libretti di tutti gli istituti raccoglitori, comprese le casse postali. Ma il tracciato ad occhio può imputarsi di facile arbitrio; meno arbitrario parrà in ogni caso un vero e proprio procedimento d' interpolazione. Se facessimo passare una curva per 3 punti convenientemente scelti, ad es., per i punti corrispondenti ai logaritmi del numero dei libretti superiori rispettivamente a 1 lira, a 2000 lire e a 4000 lire, avremmo:

$$\log y_{I} = A - \alpha (\log 1)^{n} = 6.65967$$
$$\log y_{II} = A - \alpha (\log 2000)^{n} = 5.31860$$
$$\log y_{III} = A - \alpha (\log 4000)^{n} = 4.87402$$

donde l' equazione:

$$\log y = 6.65967 - 0{,}02669 (\log x)^{3.28}.$$

Da questa formola si ricaverebbe una seriazione di tutti i libretti (e per differenza poi quella dei libretti postali) già molto verosimile, ma ancora con errore in eccesso. Vale a dire, applicando alle singole categorie così calcolate di libretti postali i convenienti valori medii, si otterrebbe un totale superiore ai 400 milioni ricordati. Ad ogni modo il risultato sarebbe sempre più attendibile di quello che può aversi dalla ipotesi di proporzionalità.

L' interpolazione della funzione $(\log x)^3$ risponde meglio al nostro scopo, sebbene dia luogo a divergenze di qualche rilievo tra i dati teorici e quelli calcolati per i libretti superiori a 2000 lire. Procedendo con metodo analogo a quello seguito a pag. 190, si ha:

x	$(\log x)^3$	$\log y$	$\delta (\log x)^3$	$\delta \log y$
1.	0	6.65967	— 35.07145	+ 1.33724
2000.	35.97069	5.31860	+ 0.89924	— 0.00383
3000.	42.03968	5.06864	+ 6.96823	— 0.25379
4000.	46.73617	4.87402	+ 11.66472	— 0.44841
5000.	50.61071	4.69124	+ 15.53926	— 0.63119
Medie . . .	35.07145	5.32243	∓ 35.07145	± 1.33723

$$1.33723 : -35.07145 = -0{,}038129$$

ossia:

$$\log y - 5.32243 = -0{,}038129 ([\log x]^3 - 35{,}07145),$$
$$\log y = 6.65967 - 0{,}038129 (\log x)^3$$

Applicando la qual formola, si ottengono questi risultati:

	Libretti di tutti gli Istituti secondo la formola	Libretti delle Casse non postali	Differenza = Libretti delle Casse postali
Sopra 1 lira	4.567.433	1.894.306	2.673.127
» 50 lire	2.969.550	1.283.264	1.686.286
» 100 »	2.262.760	1.078.240	1.184.520
» 500 »	812.905	626.709	186.196
» 1000 »	426.766	407.741	19.025
» 2000 »	194.151	208.256	— **14.105** (Divergenze tra il calcolo e l'osservazione)
» 3000 »	113.957	117.124	— **3.167**
» 4000 »	75.452	74.821	+ **631**
» 5000 »	53.696	49.118	+ **4.578**

Procedendo per classi di valore, i libretti postali risulterebbero:

	Numero dei libretti postali
Tra 1 e 50 lire	986.841
» 50 e 100 »	501.766
» 100 e 500 »	998.324
» 500 e 1000 »	167.171
Oltre 1000 lire	19.025
	2.673.127

Tenuto conto del rapido decrescere della serie, i valori medii probabili da applicarsi alle singole categorie sarebbero in cifre tonde: 20, 70, 220, 610 e 1200; adottandoli come moltiplicatori e sommando i risultati, si arriva appunto alla cifra voluta di 400 milioni.

L'ipotesi della *continuità* può dunque ritenersi reale.

Per motivi di chiarezza non abbiamo voluto complicare il procedimento con altre considerazioni. Tra l'altre se avessimo potuto mettere in conto le operazioni di risparmio, che si fanno da alcuni istituti qui non contemplati (Casse di prestanza agrarie, Opere pie, ecc.), la formola lievemente modificata avrebbe dato, per i depositi superiori a 2000 lire, ancor minori divergenze tra il calcolo e l'osservazione, e l'ipotesi della continuità della curva avrebbe corrisposto anche meglio alle esigenze di una statistica congetturale.

§ **6.** Esiste una legge di correlazione tra redditi e patrimoni? Quale patrimonio ha probabilità di trasmettere a' suoi eredi un capofamiglia, che disponga di un reddito annuo, mettiamo, di 5000 lire; e quale patrimonio invece ha probabilità di trasmettere ai suoi eredi quel capofamiglia, che dispone di un reddito annuo di 10.000 lire? Naturalmente si tratta di individui medii o tipici. L'osservazione comune

non può dirci altro se non che a redditi piccoli corrispondono, in generale, piccoli patrimoni; a redditi grandi, patrimoni grandi. Col crescere infatti del reddito lordo e netto, cresce la parte che s'investe in beni durevoli di consumo (oggetti d'ornamento, mobili, case, ville) e la parte capitalizzabile in scorte industriali, macchine, titoli, ecc. Un piccolo patrimonio addossa il cómpito della formazione del reddito quasi per intero all'attività personale del possessore; un grande patrimonio sostiene spesso da solo il cómpito di formare il reddito, senza richiedere il concorso del lavoro diretto di chi lo possiede. L'osservazione comune suggerisce pure che in un paese l'entità dei patrimoni privati può variare da zero a molti milioni, fino al miliardo, mentre per i redditi annui nè si discende a zero (ogni individuo dovendo pur vivere su un reddito di qualunque fonte, fosse anco quella della carità privata o pubblica, del furto, ecc.), nè si va, di regola, oltre il limite segnato dall'interesse al quale possiamo supporre investito il maggiore tra i patrimoni privati esistenti. Il possessore di un miliardo di patrimonio non avrà, verosimilmente, più di 50 o 60 milioni di reddito annuale. Quindi è facile immaginare che in un paese la progressione dei redditi, dal più piccolo al più grande, deve avere uno sviluppo più breve della progressione dei patrimoni privati.

Le statistiche, sebbene affatto rudimentali in questo àmbito di ricerche, ci permettono di formulare alcune leggi empiriche:

1) *Se i redditi crescono in progressione geometrica per 2, i patrimoni corrispondenti crescono*, a un bel circa, *in progressione geometrica per* 3; o, in altre parole, a un reddito doppio di un altro corrisponde un patrimonio non doppio, ma triplo di quello che corrisponde al secondo.

2) *Le professioni accessibili agli individui forniti di patrimoni inegualmente elevati, sono tanto più rimunerative, quanto più elevato il patrimonio.* In pari tempo però, col crescere del patrimonio, cresce il numero degli individui, che potendo vivere del puro reddito patrimoniale, si ritengono esonerati dall'obbligo di esercitare una professione a scopo di lucro.

3) Conseguenza d'indole finanziaria della prima legge empirica è che *una imposta proporzionale sul patrimonio funziona come se fosse un'imposta progressiva sul reddito;* e viceversa: un'imposta stabilita in misura proporzionale ai redditi apparirebbe men che proporzionale nel ragguaglio ai patrimoni corrispondenti.

Il seguente prospetto risulta da elementi forniti, per i redditi, dalla Statistica delle tasse comunali applicate negli anni 1881-1884, e per i patrimoni, dalle Statistiche delle successioni per gli esercizi 1890-1891, 1892-1893, 1900-1901, 1901-1902, 1902-1903 (1).

---

(1) Nel 1881-1884 ben 4700 Comuni applicavano la tassa di famiglia e i redditi esenti e tassati erano 2.900.000 su 14.700.000 abitanti; il che fa ritenere che

Categorie di redditi totali	Numero dei capifamiglia esistenti	Categorie di patrimoni	Numero dei capifamiglia autori di successioni	Numero dei capifamiglia esistenti (la precedente colonna moltiplicata per 36)
Meno di 1000 lire . .	4.453.566	Meno di 4000 lire	130.403	4.694.508
Da 1.000 a 2.000 lire	702.117	Da 4.000 a 10.000 lire	13.854	498.744
» 2.000 a 4.000 »	314.780	» 10.000 a 50.000 »	9.456	340.416
» 4.000 a 6.000 »	76.442	» 50.000 a 100.000 »	1.552	55.872
» 6.000 a 10.000 »	51.202	» 100.000 a 300.000 »	954	34.344
» 10.000 a 20.000 »	26.670	» 300.000 a 500.000 »	171	6.156
» 20.000 a 50.000 »	8.818	» 500.000 a 1 milione »	100	3.600
Oltre 50.000 lire . . .	1.989	Oltre 1 milione di lire	54	1.944
	5.635.584		156.544	5.635.584

Ora, se si ammette che ad ogni reddito totale di famiglia corrisponde nel corso di una generazione un patrimonio ereditario e, nell'istante considerato, un patrimonio della famiglia vivente, e se si ammette che in generale ai redditi più piccoli corrispondono i più piccoli patrimoni, ai redditi maggiori, i patrimoni maggiori, ai medii i medii, la questione di sapere quale ammontare di patrimonio è concomitante di un dato ammontare di reddito, sarà presto risolta col confronto della seconda e dell'ultima colonna del prospetto. Per esempio:

---

in tutto il Regno si avessero 5.635.584 redditi di capifamiglia. Però la classificazione per valore fu indicata solo per 19 capoluoghi di provincia, per 46 capoluoghi di circondario e per 355 Comuni non capoluoghi. Su questi elementi parziali, tenuto conto dei coefficienti d'importanza dei diversi gruppi di Comuni, in ragione della rispettiva popolazione, fu calcolata la probabile classificazione dei redditi di famiglia per tutto il Regno a quella data. Se ammettiamo — ipotesi necessaria e legittima — che ogni reddito di famiglia dia luogo nel corso di una generazione (36 anni) ad una trasmissione ereditaria di patrimonio, il numero medio delle successioni che si aprono ogni anno per morte di capifamiglia dovrebb'essere di 156.544 (= 5.635.584 : 36). Le successioni dei cinque anni finanziari indicati nel testo provengono certo quasi tutte da capifamiglia contemplati nella statistica dei redditi per il periodo 1881-1884; quindi manteniamo invariato il totale annuo di 156.544 successioni. Ora, moltiplicando per 36 il numero dei patrimoni ereditarî compresi nelle diverse classi di valore, avremo il numero approssimativo dei patrimoni dei capifamiglia viventi (ultima colonna del prospetto). Vero è che le nostre statistiche successorie non distinguono tra capifamiglia e membri di famiglia, ma vi sono ragioni per credere che i primi formino la gran maggioranza e che anzi, a partire dal limite di 4 o 5 mila lire, costituiscano la quasi totalità degli autori di successioni.

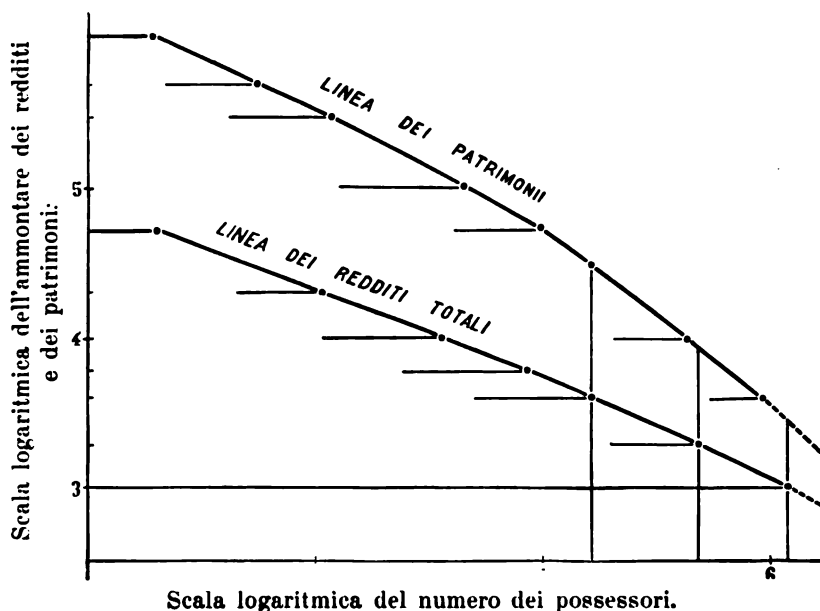
I redditi da 2000 lire in su, come facilmente si ricava dalla seconda colonna, sono 479.901. Da qual limite di valore bisogna partire per avere giusto altrettanti patrimoni? L'interpolazione indica il limite di 9016 lire. Quindi diremo: i patrimoni da 9016 lire in su sono tanti quanti i redditi da 2000 lire in su; e per le ipotesi fatte debbono provenire da questi e non da redditi inferiori.

I redditi eccedenti 4000 lire sono 165.121. Da qual limite di valore si hanno giusto altrettanti patrimoni? L'interpolazione dà il limite di 29.170 lire. Dunque, per le ipotesi concesse, i patrimoni da 29.170 lire in su provengono o si accompagnano a redditi da 4000 lire in su, non già a redditi inferiori a questa somma.

Procedendo in questo modo, e interpolando, ove occorra, anche la serie dei redditi, si ottengono i risultati seguenti:

Redditi (lire)	Patrimoni corrispondenti (lire)	Redditi (lire)	Patrimoni corrispondenti (lire)
1.000	2.850	5.000	43.150
2.000	9.016	10.000	117.500
4.000	29.170	20.000	317.700
8.000	85.700	40.000	750.000
16.000	233.300		
32.000	572.800		

Le quali serie ci risultano non da un calcolo vero e proprio, ma dalla semplice lettura di un diagramma a doppia scala logaritmica e di dimensioni piuttosto grandi, diagramma che qui riproduciamo in dimensioni molto ridotte:

Scala logaritmica del numero dei possessori.

La figura è disegnata nel solito modo, integrando le seriazioni e sostituendo ai numeri naturali i loro logaritmi. Per determinare il limite di valore, oltre il quale si hanno tanti patrimoni quanti sono i redditi superiori ad un cert'altro limite dato, basta evidentemente far passare per l'estremo dell'ascissa del reddito un'ordinata, prolungandola fino all'incontro della curva dei patrimoni. L'altezza totale di questa ordinata dà, in logaritmi, il limite cercato. Per non complicare la figura tracciammo solo le ordinate passanti per gli estremi delle ascisse dei possessori di redditi oltre 1000 lire, oltre 2000 e oltre 4000.

Uno sguardo ai risultati, cui siamo pervenuti, conferma che ad un reddito doppio di un altro corrisponde un patrimonio all'incirca triplo. La regola non è assoluta; per redditi piccoli si ha una progressione anche più che tripla dei patrimoni corrispondenti; per redditi elevati, una progressione men che tripla. Ciò dipende in parte dalla circostanza che le successioni di milionari e di semimilionari in Italia si son fatte singolarmente scarse negli ultimi anni, cosa spiegabile come un riflesso del recente e rapido sviluppo della ricchezza mobiliare, la quale sfugge più facilmente al fisco nelle successioni. Comunque sia, la norma empirica così come fu formulata può bastare per un primo approccio alla vera legge del fenomeno. In calcoli di questa specie è illusorio inseguire una maggior precisione formale; l'osservazione statistica, se si farà, costituirà il miglior correttivo degli enunciati, che pel momento non possono avere altra base che nelle ipotesi.

La seconda conclusione, che si ricava da questa correlazione tra redditi e patrimoni, è che le professioni, rese accessibili all'individuo dal possesso di una certa ricchezza, sono tanto più rimunerative quanto maggiore il patrimonio a disposizione, che serve a sostenere le spese di un lungo e costoso tirocinio e a fornire i mezzi d'impianto dell'azienda o dell'esercizio professionale. Infatti la serie dei redditi totali, che abbiamo data or ora, può scindersi in due componenti: l'una del reddito proveniente dallo stesso patrimonio, supposto investito uniformemente nelle varie categorie al saggio, mettiamo, del 5 %; l'altra del reddito propriamente personale, dovuto al lavoro dello stesso possessore. Sicchè, per es., un reddito totale di 2000 lire, a cui si accompagna un patrimonio di 9016 lire, può ritenersi composto di 451 lire, frutto dell'investimento dei beni patrimoniali e di 1549, frutto dell'attività professionale. Calcolando in simile maniera, si avrebbe:

Reddito totale (lire)		Reddito proprio del patrimonio concomitante		Reddito del lavoro personale
1.000	=	143	+	857
2.000	=	451	+	1.549
4.000	=	1.458	+	2.542
8,000	=	4.285	+	3.715
16.000	=	11.665	+	4.335
20.000	=	15.885	+	4.115
32.000	=	28.640	+	3.360
40.000	=	37.500	+	2.500

Rileviamo subito che a partire dal limite di 16.000 lire pei redditi totali, la parte che spetterebbe al lavoro personale del possessore diminuisce; ma ciò non vuol dire che diminuisca realmente la rimunerazione delle professioni; vuol dire soltanto che molti ricchi vivono sul puro reddito patrimoniale senza esercitare una professione a scopo di lucro, e che questa loro condotta riduce la media del reddito personale calcolato sull'intera classe. La qual cosa certo si verifica, sebbene in minori proporzioni, anche presso gli strati sociali meno elevati in ricchezza. Quindi si conclude che la serie dei redditi professionali crescerebbe con rapidità anche maggiore, se in ogni categoria si tenessero fuori conto gli individui improduttivi — improduttivi nel senso che vivono sui frutti del patrimonio, senza integrarli con una professione, che pur potrebbero esercitare.

La statistica, per quanto congetturale, vien così a confermare l'osservazione comune. Nelle classi più umili della società, mancando un patrimonio sufficiente per accedere a professioni assai rimunerative, ma richiedenti un lungo e costoso tirocinio, le forze di lavoro vengono utilizzate già nell'adolescenza; l'individuo si guadagna presto la vita, ma guadagna poco, il puro necessario. Nelle classi relativamente agiate, si guadagna tardi, ma si guadagna più che in proporzione. I diversi gruppi sociali, pel fatto stesso della ricchezza maggiore o minore di cui dispongono, costituiscono, in quanto concerne la scelta delle professioni, dei gruppi non concorrenti, o tali almeno che la concorrenza riesce facile da strati superiori ad inferiori, difficile invece nel senso inverso.

L'ultima conseguenza, che ricavasi dalla correlazione empirica tra redditi totali e patrimoni, è, come dicemmo, d'indole finanziaria. Un'imposta proporzionale sui patrimoni funzionerebbe come un'imposta progressiva sui redditi totali corrispondenti. Infatti suppongasi stabilita l'aliquota fissa dell'uno per cento sui patrimoni indicati a pag. 335; l'imposta pagata dalle diverse categorie di contribuenti, ragguagliata ai loro redditi totali, rappresenterà una aliquota non costante, ma crescente:

Redditi (lire)	Patrimoni corrispondenti (lire)	Imposta proporzionale dell'1 % sui patrimoni (lire)	La stessa imposta ragguagliata al reddito equivale a
1.000	2.850	28,50	2,85 %
2.000	9.016	90,16	4,51 »
4.000	29.170	291,70	7,29 »
8.000	85.700	857,00	10,71 »
16.000	233.300	2333,00	14,58 »
32.000	572.800	5728,00	17,90 »

Questo principio, induttivamente ben stabilito, potrebbe trovare una interessante applicazione negli ordinamenti finanziarî e costituire un termine naturale di conciliazione tra i fautori dell'imposta proporzionale e quelli dell'imposta progressiva. La capacità contributiva dei cittadini è misurata non solo dal reddito, ma anche dal patrimonio. D'altronde le spese dello Stato sono in parte destinate a costituire o reintegrare il patrimonio nazionale (tali le spese per costruzioni di ferrovie, di strade ordinarie, porti, fortificazioni, navi da guerra, ecc., ecc.) e sembra logico che esse siano sostenute dai cittadini piuttosto in proporzione dei loro patrimoni, che dei loro redditi; in parte invece si fanno per l'esercizio ordinario dei pubblici servizi, senza che lascino traccia nel patrimonio; e a questa parte dovrebbero concorrere i cittadini in proporzione dei loro redditi. Fissata così l'estensione del fabbisogno finanziario nelle due forme anzidette, riuscirebbe agevole determinare le aliquote di due imposte, l'una proporzionale sui patrimoni, l'altra proporzionale sui redditi. La prima però funzionando, per ciò che si è detto, come se fosse una sovrimposta progressiva sul reddito, attuerebbe il principio della progressione, senza incorrere nell'accusa di arbitrarietà, che si suol muovere all'applicazione pratica di questo principio.

Tali le conseguenze immediate delle due ipotesi fondamentali, dalle quali abbiamo preso le mosse. Altre conseguenze importanti verrebbero dall'estendere questo genere di indagini ai fenomeni economici, graduabili per valore, i quali hanno rapporti non dubbi col reddito o col patrimonio. Gli è evidente che, salvo casi eccezionali, una persona provvista di meschino reddito e di meschino patrimonio non riuscirà ad indebitarsi per cambiali dell'importo di molte migliaia di lire; nemmeno figurerà tra i maggiori azionisti di una Banca o tra i depositanti di forti somme a risparmio o in conto corrente; nemmeno figurerà come attore o convenuto in cause civili o commerciali di gran valore; nemmeno viaggerà in prima classe sulle ferrovie o fumerà le qualità di tabacco più costose..... Ciascuno intende che queste particolari manifestazioni della vita economica sono in rapporto più o meno stretto coi mezzi che l'individuo ha a sua disposizione. Orbene, chi descrivesse col solito metodo dei diagrammi a doppia scala logaritmica la ripartizione per valore dei depositi a risparmio, delle cambiali emesse, dei protesti cambiari, dei fallimenti, delle vendite volontarie o forzate di immobili, ecc., chi descrivesse le sentenze delle varie magistrature distinte secondo il valore dell'oggetto controverso o l'addensamento della popolazione nelle abitazioni, secondo l'ammontare degli affitti, o il consumo delle varie qualità di tabacchi, secondo il loro prezzo, o il movimento dei viaggiatori di ferrovia secondo l'importanza delle classi combinata colla lunghezza del percorso, otterrebbe delle curve, che or somigliano a quella dei redditi, or somigliano a quella dei patrimoni. Vi sarà dunque una ragione di siffatte

somiglianze, e la ragione deve ricercarsi nelle particolari attinenze, in cui si trovano questi, che possiamo chiamare *fenomeni economici di sottordine*, col fenomeno più generale della ripartizione dei redditi e dei patrimoni in un paese. Quali ipotesi siano indispensabili per una soluzione approssimativa del problema, qui non diremo. Una analisi altra volta da noi tentata in questa direzione è riuscita cosa ben piccola, informe e inadeguata al bisogno; ma la miniera — indicata poichè è ben una miniera di fatti nuovi e interessanti — non potrà non invitare esploratori arditi.

§ **7.** Quando da premesse, in parte induttivamente stabilite, in parte confortate da ipotesi plausibili o reali, ricaviamo nella forma solita del sillogismo certe conseguenze, la Statistica si riserba l'ufficio modesto, ma utile, di strumento di verifica. L'accordo della ricostruzione ideale dei fenomeni colla realtà loro prova la correttezza del ragionamento.

Poniamo di avere accertato per mezzo di osservazione comune o di osservazione statistica che nelle provincie italiane già soggette al governo pontificio perdurano assai più che altrove i matrimoni semplicemente religiosi, i quali solo dopo parecchi anni di convivenza vengono regolarizzati col rito civile. Combinando questa premessa con altre minori, noi inferiamo:

1) che la natalità illegittima, a parità d'altre circostanze, deve essere maggiore in dette provincie che in altre parti d'Italia, grazie al contributo speciale che le forniscono le nozze semplicemente religiose;

2) che il carattere di cotesta natalità illegittima nelle provincie ex-pontificie dev'essere tale da farla differire ben poco dalla legittima, trattandosi di bambini allevati in seno a famiglie moralmente bene costituite, sebbene irregolari di fronte al Codice civile; e che pertanto la mortalità degli illegittimi deve quivi differire poco da quella dei legittimi; mentre la differenza è grande nelle altre parti d'Italia, dove gli illegittimi sono per lo più abbandonati e trascurati dai loro genitori;

3) che per analoghi motivi la mortalità delle madri illegittime, per malattie inerenti allo stato di fecondità, deve differire poco da quella delle legittime nelle provincie ex-pontificie; ed altrettanto dicasi della nati-mortalità dei bambini; laddove il contrasto dev'essere ed è grande nel resto del paese, dove i parti illegittimi avvengono spesso senza la necessaria assistenza;

4) che in dette provincie scarsa sarà la fecondità dei matrimoni regolari, posto che essi sono molto spesso preceduti da un periodo di convivenza feconda illegittima;

5) che l'età media degli sposi con rito civile dev'essere, a parità d'altre circostanze, più alta nelle provincie ex-pontificie, poichè ivi molte nozze civili non sono che tardive legittimazioni di convivenze durate più o meno a lungo col semplice vincolo religioso;

6) che il numero dei vedovi passanti a seconde nozze risulterà lì più scarso che altrove, perchè essendo, come si disse, più alta l'età media degli sposi al momento dell'unione civile, le vedovanze vere e proprie debbono cadere in età proporzionatamente inoltrata, cioè in età in cui è meno facile il convolare a seconde nozze. In altri termini, pochi gli sposi giovani, poche le vedovanze in giovane età, che forniscono il maggior tributo alle seconde nozze. Aggiungasi che se Tizio perde la sua compagna, a lui congiunta con solo vincolo religioso dal dì che questo perdette la sua validità legale, e passa ad altra unione, stavolta regolare, con altra donna, egli verrà considerato dall'ufficiale di stato civile non come un vedovo passante a secondo matrimonio, ma come un celibe che contrae le prime nozze;

7) che la distribuzione per mesi dei matrimoni civili nelle provincie ex-pontificie deve risentire meno che altrove l'influenza dei divieti della Chiesa nelle epoche dell'Avvento e della Quaresima; perchè, mentre nel resto d'Italia il rito civile e il religioso sogliono coincidere e quindi il primo si risente degli ostacoli che si oppongono al secondo, nelle provincie nominate la coincidenza di tempo spesso manca; il rito civile segue non di rado a più anni di distanza il religioso e rimane così svincolato dall'osservanza dei divieti ecclesiastici.

Queste ed altre deduzioni si trovano punto per punto confermate dall'osservazione statistica.

Anzitutto vediamo quanto maggiore fu la riduzione dei matrimoni regolari nelle provincie ex-pontificie in confronto delle altre italiane, dopo l'introduzione del rito civile.

La media degli anni 1863 e 1864 può assumersi come normale.

Anni o periodi	Provincie ex-pontificie (esclusa quella di Roma) — Matrimoni		Altre provincie italiane (escluse quelle di Roma e del Veneto) — Matrimoni	
	N° assoluto	N° relativo	N° assoluto	N° relativo
1863-1864 (media annua)	17.457	**100,00**	162.287	**100,00**
1865	(19.075)	...	(186.925)	...
1866	10.275	**58,86**	111.862	**68,93**
1867	11.707	**67,06**	136.440	**84,07**
1868	11.925	**68,31**	148.494	**91,50**
1869	13.138	**75,26**	169.672	**104,55**
1870	12.864	**73,69**	155.203	**95,64**
1871	12.537	**71,82**	160.040	**98,62**

Abbiamo dovuto escludere la provincia di Roma e quelle del Veneto perchè ivi il matrimonio civile non fu introdotto che dal febbraio e dal settembre 1871, rispettivamente.

Il nostro prospetto dimostra che nelle provincie ex-pontificie la depressione della nuzialità regolare fu più accentuata e persistente che altrove (1). Le conseguenze si leggono chiare in questo secondo prospetto:

Anni o periodi	Provincie ex-pontificie (esclusa quella di Roma)				Illegittimi ed esposti per 100 nati	
	Totale annuo o media annua delle nascite	Nati			Provincie ex-pontificie	Altre provincie italiane (comprese le venete anche prima del 1867)
		legittimi	illegittimi	esposti		
1863-1866	91.415	87.234	780	3401	4,57	5,03
1867	87.756	81.937	2.188	3631	6,63	5,48
1868	85.921	78.910	3.408	3603	8,16	5,82
1869	90.401	81.935	4.874	3592	9,37	5,63
1870	87.482	77.334	6.517	3631	11,60	5,89
1871	88.344	76.609	7.887	3848	13,28	5,95
1872	88.379	74.941	9.464	3974	15,20	6,08
1873	87.606	73.455	10.335	3816	16,15	6,05

Dunque: regresso nella natalità legittima delle provincie ex-pontificie; progresso nella illegittima. Di questa si hanno due categorie: gli illegittimi accertati come tali, massime per riconoscimento da parte di uno o di entrambi i genitori, e gli esposti, il cui stato civile è ignoto, ma che si presumono illegittimi. È precisamente la prima di queste due categorie, che si segnala per una rapida progressione nelle dette provincie. Nelle altre parti d'Italia, come fu debole e poco persistente la riduzione della nuzialità regolare, così riuscì poco accentuato l'incremento della natalità illegittima; anzi dopo il 1872 si ha un regresso.

Se poniamo in conto il Lazio, i cui dati si hanno dal 1872, il quadro si completa.

---

(1) Il dicembre del 1865 fu segnalato in tutte le provincie italiane d'allora da una eccezionale frequenza di nozze (forse 17 o 18 mila in più dell'ordinario), molte famiglie avendo voluto profittare dell'ultimo mese utile alla celebrazione legale del matrimonio col semplice rito religioso. Attribuendo tali nozze anticipate all'anno 1866, questo figurerebbe con 11.475 matrimoni delle provincie già pontificie e con 128.162 delle altre d'allora; mentre l'anno 1865 ridotto a 17.873 e a 170.625 matrimoni, rispettivamente, rientrerebbe nel novero degli anni normali.

Anni o periodi	Provincie ex-pontificie (compresa quella di Roma)				Illegittimi ed esposti per 100 nati	
	Totale annuo o media annua delle nascite	Nati			Provincie ex-pontificie	Altre provincie
		legittimi	illegittimi	esposti		
1872 . . .	117.990	101.943	11.232	4815	13,60	6,08
1873 . . .	115.948	98.467	12.677	4804	15,08	6,05
1874 . . .	112.223	93.768	13.727	4728	16,44	6,05
1875-1877. . .	122.473	100.766	16.675	5033	17,72	5,66
1878-1880. . .	118.496	96.215	17.219	5061	18,80	5,75
1881-1883. . .	129.169	103.623	20.520	5026	19,78	5,86
1884-1886. . .	136.669	108.679	21.893	6097	20,48	5,74
1887-1889. . .	142.712	113.383	23.190	6140	20,55	5,50
1890-1892. . .	140.923	112.732	22.401	5789	20,00	5,25
1893-1895. . .	140.259	114.377	20.500	5382	18,45	5,02
1896-1898. . .	135.654	112.170	18.592	4892	17,31	4,80
1899-1902. . .	132.746	111.889	16.319	4538	15,71	4,52

Dal 1884 in poi la categoria « illegittimi » comprende i soli illegittimi riconosciuti all' atto della nascita da uno almeno dei genitori; l' altra categoria comprende, oltre gli esposti, gli illegittimi accertati come tali dall' ufficio di stato civile, ma non riconosciuti dai genitori.

Così, mentre nel resto d'Italia si torna presto all'antica quota normale di natalità irregolare, nelle provincie già pontificie il crescendo continua fino al 1887-89, e dopo d'allora soltanto la curva ripiega in basso, segno di vittoria contrastata, ma sicura del rito civile anche presso coteste popolazioni. Senza ricorrere per la dimostrazione alla statistica dei matrimoni civili, chiederemo la prova alle legittimazioni di figli naturali per susseguente matrimonio o per decreto reale. La frequenza di codeste legittimazioni nelle provincie già pontificie è grandissima, in ragione di popolazione; ma è anche grande per rapporto alle nascite illegittime. Infatti:

PERIODI	Nascite illegittime		Periodi	Legittimazioni		Legittimati per 100 nati illegittimi	
	Provincie già pontificie	Altre provincie		Provincie già pontificie	Altre provincie	Provincie già pontificie	Altre provincie
1884-1886 (media annua)	27,990	56.117	1885-1887	5.432	8.430	19,4	15,0
1890-1892 » »	28,190	50.758	1891-1893	10.622	10.717	37,7	21,1
1893-1895 » »	25,882	48.564	1894-1896	11.611	12.583	44,9	25,9

Abbiamo riferito le legittimazioni di ogni periodo alle nascite illegittime di un periodo anteriore di un anno, perchè appunto la regolarizzazione delle unioni col rito civile ha luogo di solito ad intervallo di almeno un anno dalla nascita dei figli naturali.

Una delle conseguenze logiche sopra affermate era che la natimortalità degli illegittimi deve essere poco differente da quella dei legittimi nelle provincie già pontificie, mentre è molto diversa nelle altre parti del Regno. Ecco i dati dell'osservazione pel sessennio 1884-89 (medie annuali):

PROVINCIE	Legittimi			Illegittimi ed esposti		
	Nati-vivi e nati-morti	Nati-morti	Percentuale dei nati-morti	Nati-vivi e nati-morti	Nati-morti	Percentuale dei nati-morti
Già pontificie. .	115.328	4.297	3,73	30.015	1355	4,51
Altre del Regno	964.959	32.607	3,38	59.483	3970	6,67

La differenza di nati-mortalità nelle due categorie, che si limita a 0,78 punti per le provincie già pontificie, si eleva a 3,29 punti per il resto del Regno. La prova è concludente.

Altra conseguenza affermata era che la mortalità degli illegittimi nei primi anni di età dovesse differire poco da quella dei legittimi nel gruppo di provincie considerato, mentre differisce notevolmente nel resto del Regno. Pel primo anno d'età le osservazioni del sessennio 1885-90 dànno questi risultati:

PROVINCIE	Legittimi			Illegittimi ed esposti		
	Semisomma dei nati-vivi nel 1884-1889 e nel 1885-1890	Morti in età da 0 a 1 anno nel 1885-1890	Quoziente di mortalità	Semisomma dei nati-vivi nel 1884-1889 e nel 1885-1890	Morti in età da 0 a 1 anno nel 1885-1890	Quoziente di mortalità
	Medie annue.			Medie annue.		
Già pontificie. .	111.217	21.350	19,2 %	28.692	6.426	22,4 %
Altre del Regno	928.764	175.096	18,9 %	54.913	15.260	27,8 %

La prova è raggiunta. I bambini legittimi nel corso del primo anno d'età soccombono in ragione del 19,2 % nella zona speciale da noi considerata; e gli illegittimi in una ragione poco superiore, 22,4 %. Nel restante del Regno le quote sono 18,9 e 27,8 %, con una differenza assai più grave a carico degli illegittimi.

Veniamo alla mortalità delle madri per malattie inerenti allo stato di fecondità. Le statistiche del periodo 1888-92 non ci permettono di separare i dati delle provincie romagnole da quelli dell'Emilia superiore; quindi per chiarezza di confronti procederemo così:

COMPARTIMENTI	Nati legittimi	Coniugate morte per malattie di gravidanza, parto e puerperio	Coniugate morte per 1000 nati	Nati illegittimi ed esposti	Nubili e vedove morte per malattie di gravidanza, parto e puerperio	Nubili e vedove morte per 1000 nati
Marche, Umbria, Lazio	78.461	382,4	**4,87**	17.886	85	**4,75**
Emilia	70.420	343,8	**4,88**	13.355	62,8	**4,70**
Altri compartimenti del Regno .	889.303	3857	**4.34**	49.490	307,6	**6,22**

Madri legittime e madri illegittime soccombono in quasi egual proporzione nelle Marche, Umbria e Lazio, nonchè nell'Emilia, che per più di metà fa parte dello speciale territorio demografico in esame; anzi sembra persino che la condizione di illegittimità attenui quivi invece di aggravare i pericoli inerenti allo stato di fecondità. Nei restanti compartimenti del Regno invece la differenza di condizioni appare rilevante; dalla quota di 4,34 madri legittime morte per 1000 puerpere si passa a quella di 6,22 madri illegittime.

La fecondità dei matrimoni presenta questo divario:

PROVINCIE	Matrimoni (media annua del 1883-1888)	Nati-vivi legittimi (media annua del 1884-1889)	Nati-vivi per ogni matrimonio
Già pontificie . . .	28.092	111.031	**3,95** (1)
Altre provincie . .	207.109	932.352	**4,50**

A voler tener conto anche dei nati-morti legittimi, le quote della fecondità salirebbero, rispettivamente, a 4,11 e 4,66.

L'alta proporzione delle prime nozze nelle provincie già pontificie è confermata dai seguenti dati riferibili al periodo 1883-88 (media annua):

MATRIMONI	Provincie		Percentuali	
	Già pontificie	Altre provincie	Provincie già pontificie	Altre provincie
Tra celibi e nubili . .	24.900	174.711	88,64	84,36
» celibi e vedove .	691	6.828	2,46	3.29
» vedovi e nubili .	2.020	18.056	7,19	8,72
» vedovi e vedove .	481	7.514	1,71	3,63
Totali . . .	28.092	207.109	100	100

(1) Chi risalisse colle indagini ai primi anni del Regno troverebbe per le

Avanti il 1865 la proporzione delle prime nozze nel territorio già pontificio (esclusa Roma) non differiva gran fatto da quella osservata nel restante Regno.

La scarsità dei vedovi passanti a seconde nozze parrebbe dovesse tenere bassa l'età media degli sposi nelle provincie già pontificie; ma per altro motivo da noi detto, e che è di portata assai maggiore, questa età media riesce invece più alta che altrove. Limitiamoci a considerare la ripartizione delle spose per età, secondo le medie del periodo 1883-1886.

CLASSI D'ETÀ	Provincie a nord delle ex-pontificie		Provincie ex-pontificie		Provincie a sud delle ex-pontificie (comprese le isole)	
	Cifre assolute	%	Cifre assolute	%	Cifre assolute	%
Meno di 19 anni . . .	6.804	6,33	873	3,16	11.811	11,87
19-20 anni	17.089	15,89	2428	8,79	14.738	14,81
21-22 »	**23.667**	**22,01**	4608	16,69	**17.629**	**17,72**
23-24 »	19.189	17,84	**4935**	**17,87**	14.867	14,94
25-26 »	12.642	11,75	4103	14,86	11.182	11,24
27-30 »	13.152	12,23	4874	17,65	12.571	12,63
31-40 »	10.732	9,98	4452	16,12	10.952	11,01
41-50 »	3.100	2,88	1068	3,87	3.930	3,95
Oltre 50 anni	1.138	1,06	267	0,97	1.759	1,76
Età ignota	33	0,03	9	0,03	74	0,07
Totali . . .	107.546	100,00	27.616	100,00	99,513	100,00

Le spose fino a 23 anni rappresentano nelle provincie già pontificie la quota di 28,64 %, nelle altre provincie a nord e a sud invece la quota di 44,23 e 44,40 rispettivamente. L'età media di tutte le spose appare nelle prime di due anni più elevata che nelle seconde. Questa differenza non esisteva avanti l'introduzione del matrimonio civile, come sarebbe facile provare.

In conseguenza dell'introduzione del rito civile la distribuzione per mesi dei matrimoni si è così modificata nelle provincie già pontificie e nel resto del Regno:

provincie già pontificie (esclusa Roma) una fecondità di quasi 5 nati per matrimonio. Ai 17.457 matrimoni del 1863-1864 (media annua) fanno riscontro appunto 87.234 nati legittimi del 1863-1866 (media annua). La quota di 5 nati, o meglio quella di 4,50 che adesso devesi considerare normale, si potrebbe ancor ricostituire oggi aggiungendo ai nati legittimi i bambini legittimati per susseguenti nozze civili e quelli premorti alla regolarizzazione dell'unione dei loro genitori.

MESI	Provincie già pontificie (esclusa quella di Roma nelle prime due serie)			Altre provincie del Regno (comprese quelle venete, anche nel 1863-1864)		
	1863-1864	1867-1869	1887-1889	1863-1864	1867-1869	1887-1889
Gennaio . . .	78,5	70,4	76,9	124,4	94,7	104,5
Febbraio . . .	108,6	104,3	108,7	133,0	139,3	128,0
Marzo	**17,1**	**79,9**	**79,1**	**36,4**	**66,4**	**72,7**
Aprile	**128,1**	**84,6**	**96,5**	**103,8**	**84,3**	**84,4**
Maggio . . .	46,4	76,1	78,4	70,7	72,8	75,8
Giugno . . .	100,2	74,2	76,1	65,6	62,0	64,1
Luglio	44,1	60,4	56,1	53,7	54,7	56,2
Agosto . . .	55,0	65,2	60,6	62,9	62,2	59,9
Settembre . .	75,6	68,6	73,9	77,6	70,2	74,2
Ottobre . . .	160,9	103,4	103,6	87,8	84,7	84,9
Novembre . .	**159,6**	**111,0**	**99,4**	**127,4**	**126,8**	**104,6**
Dicembre . .	**25,9**	**102,0**	**90,8**	**56,6**	**81,9**	**90,6**
Totali . .	1000,0	1000,0	1000,0	1000,0	1000,0	1000,0

La deficienza di nozze nel marzo, il mese della Quaresima, indicata dalla quota di 17,1 ‰ nel 1863-64 per le provincie già pontificie, e da quella di 36,4 per il rimanente del Regno, appare in gran parte colmata nei due periodi successivi scelti pel confronto; e a colmarla contribuisce specialmente l'aprile, le cui quote discendono ad un livello poco più che medio. Del pari la deficienza di nozze nel dicembre, il mese dell'Avvento, si ripara con un contingente sottratto al novembre nelle provincie già pontificie, e al gennaio nelle altre. In quelle poi meglio che in queste si accentua la tendenza ad una ripartizione regolare dei matrimoni nel corso dell'anno.

La prova è così raggiunta su tutti i punti indicati; se lo spazio nol vietasse, altre conclusioni possibili per via deduttiva verrebbero confermate dall'osservazione statistica. La qual cosa dimostra al tempo stesso la giustezza del ragionamento e la veridicità delle statistiche demografiche italiane.

§ **8.** È perfettamente conforme alla natura del metodo induttivo, che si conceda parte più o meno larga alle ipotesi. La regola dei casi osservati non si può estendere ai casi non osservati, senza un ponte di passaggio tra il noto e l'ignoto. Questo ponte è l'ipotesi. Il più semplice e generale tra gli enunciati è quello dell'uniformità del corso delle cose in natura. Ciò che è avvenuto in date condizioni, avverrà ancora, se le condizioni si ripresenteranno identiche. Nei fatti collettivi, data la grande complessità delle cause, il ripresentarsi di condi-

zioni *identiche* devesi ritenere impossibile; ma è già qualche cosa, se possiamo parlare di condizioni *simili*. A maggior ragione quindi occorre l'ipotesi come ponte di passaggio per concludere da casi osservati a casi non osservati. I casi osservati e i non osservati della stessa specie formano in realtà dei gruppi scelti, cioè in qualche modo diversamente qualificati; ma quante volte ci sia motivo di credere irrilevanti o poco rilevanti le differenze di qualificazione, le proprietà del primo gruppo si potranno riconoscere come proprietà del secondo. I criterii di proporzionalità, di continuità, ecc., trovano allora applicazione. Che poi le ipotesi debbano essere prudenti, informate al precetto del minimo arbitrio, confortate in parte da osservazione comune o storica, è cosa che per la sua evidenza si sottintende. Ma è anche evidente che per far senza delle ipotesi, bisognerebbe che l'osservazione non avesse limiti; fin che ci saranno limiti di tempo, di spazio, di precisione, di specificazione, ecc., ai rilievi statistici, le lacune intermedie o le zone vuote circostanti non si potranno colmare se non con quel mezzo. Ogni progresso dell'osservazione rende superflua una ipotesi, che dianzi appariva necessaria; ma apre la via a formularne di nuove per rischiarare un po' dell'ignoto che seguita a circondare il noto. Nessuno può così assegnare limiti all'osservazione, come non può assegnarne alle ipotesi; perchè ciò che è ignoto è infinito e ciò che è noto è e sarà sempre una quantità finita.

Appendice al Libro II.

CENNO STORICO INTORNO ALLA STATISTICA

Sommario: § 1. Le operazioni statistiche. — § 2. La letteratura statistica.

§ **1.** La conoscenza di fenomeni collettivi, specialmente dell'ordine sociale, supponendo coordinazione di ricerche e disponibilità di grandi mezzi, non esclusi i mezzi coercitivi, è venuta di buon'ora a proporsi tra i fini delle società politicamente organizzate. L'interesse amministrativo, finanziario, militare, decise presto i Governi a rilevare la grandezza del territorio, il numero degli abitanti (in ispecie quello degli atti alle armi), le fonti di ricchezza, alle quali attingere i tributi, ecc. Queste osservazioni di carattere statistico ebbero innanzi tutto scopi pratici; l'interesse scientifico non nacque se non quando l'attenzione fu richiamata dalla singolare costanza di certi fatti o dei loro rapporti numerici, e dalla conformità di alcune salienti manifestazioni della vita sociale ad un disegno, secondo alcuni, provvidenziale, secondo altri, semplicemente naturale-meccanico.

Adunque, osservazioni statistiche in servizio della pubblica amministrazione si dovettero avere anche nella remota antichità. Senza citare lontanissimi ricordi cinesi, troviamo nei libri sacri del popolo ebreo menzione di censimenti, spartizioni di terre e genealogie, le quali ultime paiono accennare ad un ordinamento dello stato civile. Gli eruditi ci informano pure delle registrazioni d'interesse pubblico, che dovevano essere in gran favore presso gli Egizii, gli Assiri e i Persiani. In Grecia la classificazione censuaria dei cittadini, l'iscrizione dei nati sui registri delle Tribù, le oblazioni religiose per ogni nascita o morte, lasciano credere ad una organizzazione, sia pure embrionale, della statistica demografica. Questa si ritrova più chiaramente attuata in Roma, dove il censo, posto a base della costituzione politica da Servio Tullio, almeno secondo la tradizione, si ripete ogni anno nel periodo repubblicano, ogni dieci e poi ogni quindici nel periodo imperiale. È forse dai censimenti che si traevano le tavole dei ***juniores*** e dei ***seniores*** in servizio della milizia. Ogni nascita, ogni morte, ogni ingresso nell'età virile si accompagnavano a determinate offerte votive; non è certo però che i dati fossero raccolti e utilizzati. Sotto gli Antonini la denunzia delle nascite al prefetto dell'erario in Roma e ai ***tabularii*** nelle provincie divenne obbligatoria.

Nel medio evo abbondano gli inventarii di beni demaniali; le corporazioni religiose, e indi le laiche, tengono registri delle loro proprietà; il clero compila liste speciali per gli atti connessi col movimento della popolazione (battesimi, nozze, funerali), liste di cui poco o nulla ci è rimasto, ma che certo formarono la base dei registri diocesani introdotti nel secolo XVI. Documento statistico-amministrativo di notevole importanza è il *liber censualis* (*Doomsdaybook*, libro del giorno del giudizio, così chiamato dal popolo sassone, vinto e spogliato) di Guglielmo il Conquistatore (1086), vero censimento dei vassalli e dei liberi e catasto completo dei terreni, delle miniere, peschiere, ecc., in Inghilterra. Alla fine del secolo XIII Venezia prescrive ai proprii agenti all'estero di presentare al Senato relazioni su quanto d'interessante poteva esserci nei paesi, presso i quali erano accreditati, così dal punto di vista politico come da quello commerciale. Nel secolo XIV, se non prima, cominciano i censimenti della repubblica. In altre città italiane, massime a Firenze, troviamo indizi di rilievi statistici.

Colla diffusione della stampa, colla formazione di grandi Stati unitari, colle scoperte geografiche, ecc., il bisogno e la possibilità dell'osservazione e dell'analisi nei fatti politici e sociali, si accrebbero. L'importanza assunta dalla politica estera, le frequenti alleanze, guerre e paci, l'istituzione degli eserciti permanenti, richiamano la necessità di indagini sulle forze dello Stato e su quelle dei rivali e degli alleati; il protezionismo con le inevitabili questioni dei trattati di commercio e colle guerre di tariffe, allega, per giustificarsi, i prospetti del movimento commerciale, accuratamente specificati nella quantità e qualità delle mercanzie entrate od uscite dal paese; il sistema monetario e quello delle imposte, fattisi meno mutabili e farraginosi che nel medio evo, semplificano molte indagini; il sorgere delle banche di emissione, il moltiplicarsi dei prestiti pubblici sotto nuove forme apparecchia materiali speciali per lo studio dei fenomeni del credito. Spesseggiano i censimenti di città o di parti del territorio, o ristretti a certe classi di popolazione; la Svezia inizia dal 1748 la regolare pubblicazione di tavole demografiche complete. L'ordine introdotto nei registri ecclesiastici, facendo meglio conoscere i fenomeni di movimento della popolazione, dà un impulso decisivo al sorgere della statistica scientifica.

Ma la vera sistemazione in grande della statistica ufficiale si può dire vanto del secolo XIX. Essa si inaugura in Francia coll'ufficio statistico del 1800, sul modello del quale organizza il suo il primo Regno d'Italia (1805). In Russia nel 1802, in Prussia nel 1805, in Austria nel 1810, in Belgio e Olanda nel 1826, ecc., si costituiscono uffici del genere, con intenti più o meno larghi. L'avversione degli stessi Governi assoluti alla pubblicità dei rilievi diminuisce gradualmente. I censimenti generali della popolazione entrano nel novero delle normali operazioni amministrative, acquistando carattere di regolare

periodicità. Parallelamente all'attività delle pubbliche amministrazioni si svolge quella delle società private di Statistica (ricordiamo la Società statistica di Parigi, sôrta nel 1829, e quella di Londra del 1834) e dei Congressi iniziati a Bruxelles nel 1833, per merito soprattutto del Quételet, coll'intento di fissare criterii più uniformi d'indagine e rendere comparabili le statistiche dei diversi Stati. L'opera dei Congressi fu sostituita in certo modo ed allargata, dopo il 1886, dall'Istituto internazionale di statistica, associazione privata, di cui fan parte direttori di uffici statistici, scienziati e parlamentari, benemeriti dei progressi degli studi statistici nei vari paesi. Le sessioni dell'Istituto, incominciate a Roma, hanno luogo regolarmente ogni due anni; la più recente, la decima, fu tenuta a Londra (1905).

§ **2.** Se dalle operazioni statistiche passiamo alla letteratura statistica, non abbiam bisogno di risalire all'antichità o al medio evo, che non ci offrono più che generici accenni degli scrittori alla necessità di ben conoscere le forze e le risorse degli Stati. Solo all'alba dell'evo moderno, opere di grossa mole, ma per lo più di scarso metodo, raccolte di notizie geografiche, storiche, politiche, economiche, che ebbero una certa fortuna libraria, compaiono in forma di *Descrizioni degli Stati*. Notevoli tra le italiane quelle del Piccolomini, del Sansovino, del Guicciardini, del Botero, ecc., tra le olandesi quella del De Laet, e per la Francia quella del D'Avity. I tedeschi poi, col loro talento sistematico, fecero di queste Descrizioni una disciplina meglio distinta dalla geografia, dalla storia e dal diritto pubblico, assegnandole per oggetto le condizioni attuali degli Stati e i fattori più importanti della loro organizzazione. Di tale indirizzo il merito spetta ad Ermanno Conring (1600-1681), seguito da molti professori universitarii e da autori di compendii della nuova scienza di Stato, tra i quali eccelle l'Achenwall (1719-1772), professore a Gottinga. L'Achenwall fa della « statistica » (denominazione che egli usa spesso, derivandola dall'italiano « *statista*, uomo di Stato » e che poi si generalizzò) una disciplina *storica* dello Stato in opposizione alla disciplina *politico-filosofica*, una base della politica pratica per via di una descrizione metodica delle cose di attualità, degne di nota in uno Stato. È in fondo lo stesso concetto del Conring, svolto un po' meno scolasticamente; è la stessa scienza descrittiva, ma fatta più autonoma dal diritto pubblico, dalla storia e geografia e dalla politica. La diffusione delle idee di Achenwall fu grande in Germania; nè le nocquero i contrasti vivaci tra coloro che dando maggior importanza ai fattori materiali, esprimibili in cifre, avrebbero voluto ridurre la nuova scienza tutta a tabelle numeriche, e coloro che considerando di preferenza i momenti formali, i fattori ideali dello Stato, traducibili solo nel linguaggio ordinario, si atteggiavano a puri continuatori del maestro. A poco a poco le due correnti si confusero in una.

Ben diverso indirizzo nella seconda metà del secolo XVII s'era iniziato in Inghilterra per opera del Graunt (1662), del Petty (1683) e dell'Halley (1693), l'indirizzo detto degli *aritmetici politici*, i quali, in luogo di descrivere le cose di attualità notevoli degli Stati, ricercavano leggi empiriche di fatti sociali. Il Graunt, compulsando i registri municipali di Londra, mise in chiaro l'equilibrio dei sessi nella popolazione, l'eccesso costante delle nascite maschili sulle femminili, l'influenza delle stagioni sui suicidii; sono quelle prime regolarità demografiche che ispireranno al Süssmilch l'idea di un ordine divino nelle cose umane. All'Halley dobbiamo il metodo ingegnoso, da noi già esposto, per la costruzione di una tavola di mortalità col solo materiale di una classificazione dei morti per età. In mancanza di veri e proprii censimenti, il problema di calcolare la popolazione in base ai registri dei decessi, occupò, dopo di lui, molti ingegni; un posto distinto in questo ramo di statistica calcolatrice spetta all'olandese Kersseboom (1738).

Ma l'intento degli aritmetici politici era innanzi tutto pratico; il materiale statistico utilizzavasi quasi unicamente per i calcoli di assicurazione sulla vita. Questi scopi pratici sono oltrepassati nell'opera del berlinese Giampietro Süssmilch (1707-1767), come sono oltrepassati quelli della statistica descrittiva di Achenwall. Süssmilch riprende su scala molto maggiore il disegno del Graunt, movendo da un punto di vista chiaramente indicato nel titolo stesso dell'opera (1741): « L'*ordine divino* nelle mutazioni del genere umano », provato dal confronto fra i nati e i morti, tra i coniugati e i nati e specialmente dal rapporto tra i maschi e le femmine nelle nascite. I dati statistici sono dunque per lui un mezzo di prova di un ordine provvidenziale ammesso *a priori* per sentimento di fede, come la *causa causarum*: il che non gli impedisce di ricercare gli antecedenti immediati e concreti dei fenomeni. Valga per tutte la diligente analisi ch'egli fa delle cause della diversa mortalità nei grandi centri di popolazione, nei piccoli centri e nelle campagne. D'altronde, tenuto conto dei tempi e della veste di pastore cristiano che Süssmilch indossava, sarebbe strano il fargli colpa dell'indirizzo teologico e teleologico dell'opera sua; mentre in capo a questa ben si sarebbero potute scrivere le parole di Newton, che Melchiorre Gioia adottò come motto per la sua *Filosofia della Statistica*: « *In hac philosophia leges deducuntur e phaenomenis et redduntur generales per inductionem* ». Dicasi piuttosto che il preconcetto dell'ordine divino esagerò un poco nel Süssmilch la tendenza a generalizzare, ad ammettere uniformità o normalità statistiche, che gli scarsi materiali di studio non provavano a sufficienza; dicasi pure che informò ad un certo ottimismo il suo pensiero intorno al problema della popolazione. Ma la statistica di Süssmilch differisce da quella di Achenwall, come una disciplina investigatrice differisce da una semplicemente descrittiva. Ed è questo l'importante. Il metodo è buono: la raccolta ed elaborazione dei dati diligente ed ingegnosa. Sì che veramente il

Süssmilch precorre la statistica moderna nel duplice aspetto che questa ha assunto, di disciplina metodologica e di scienza autonoma (demografia).

L'organizzazione della statistica ufficiale in molti Stati europei al principio del secolo XIX apparecchiava, come dicemmo, più ricco materiale agli studiosi, mentre i progressi delle scienze sperimentali e nuove vedute filosofiche portavano all'esame dei fatti sociali, realistico e non più teleologico. Le tendenze dell'epoca ebbero felice espressione nella *Fisica Sociale* e in altre opere di Adolfo Quételet (1796-1874). Il *Saggio di fisica sociale*, pubblicato nel 1835, tratta dell'uomo e dello svolgimento delle sue facoltà, con applicazioni del calcolo di probabilità e coll'evidente fine di dimostrare che i fenomeni del mondo morale soggiacciono, come i fenomeni dell'ordine fisico e biologico, a leggi determinate. Lo studio statistico dell'uomo viene così riportato entro lo studio statistico dell'ambiente in cui egli vive; e per tal via, dice bene il Wagner, la statistica diventa il legame tra le scienze naturali e le scienze storico-politiche. Il libero arbitrio, secondo lo scrittore belga, esiste bensì nell'individuo, ma la sua influenza nei fenomeni collettivi si riduce a quella di una causa accidentale, che perturba solo leggermente i risultati delle cause costanti. D'altronde l'individuo non è che una variazione del tipo, verso cui cospirano costantemente le cause o forze dell'ambiente; l'importante è di determinare « l'uomo medio ». Il concetto dell'*uomo medio*, come altri ha osservato, è in verità piuttosto instabile nella mente del Quételet; chè, alle volte, questo *uomo medio* è assunto come tipo del bello, del buono, dell'armonico in tutti i suoi caratteri; altre volte invece è un essere fittizio, pel quale tutto si verifica in conformità dei risultati medii di una collettività e fa pensare quindi a qualche cosa di mediocre in tutto, così nelle qualità fisiche, come nelle doti dello spirito. Questo secondo, ad ogni modo, è il concetto dominante. Ammesso dunque l'uomo medio come il risultato verso cui tendono le cause o forze costanti e generali dell'ambiente, i singoli individui concreti non entrano in calcolo che come deviazioni più o meno accentuate dal tipo, prodotte dall'interferenza di cause accidentali. Essi possono assimilarsi ad errori di costruzione commessi dalla natura nei tentativi che essa fa di riprodurre un modello ideato. La conferma di questo modo di vedere il Quételet credette trovarla nel fatto che un gruppo di coetanei ordinati per gradazioni di statura o di qualche altro carattere, dà luogo ad una ripartizione analoga a quella teorica degli errori di osservazione; scoperta senza dubbio interessante e feconda, ma non così concludente e generale, come l'autore entusiasta mostrava di ritenere. Maggior titolo d'onore gli fu la dimostrazione, fondata sulle prime statistiche criminali, della regolarità dei fenomeni morali; l'idea che il delinquente fosse un prodotto dell'ambiente sociale e che solo la permanenza delle cause, le quali spingono alle azioni criminose, spiegasse la costanza numerica

di questi fenomeni, rafforzava sì la tendenza determinista dell'epoca, ma richiamava pure l'attenzione del legislatore sui lati deboli della organizzazione sociale. Qualche tratto esagerato qua e là, e qualche peccato di vanità non oscurano i meriti dell'autore della *Fisica Sociale* e dell'*Antropometria*, il quale anche nel campo pratico si segnalò dirigendo il primo censimento belga e facendo la più attiva propaganda per i Congressi internazionali di statistica, allo scopo di rendere più uniformi e comparabili i rilievi statistici dei vari paesi civili.

Dopo di lui, una moltitudine di studiosi si adoprò a dilucidare i punti ancora oscuri della metodologia, ad estendere le applicazioni matematiche, a sfruttare il grande e sempre nuovo materiale, che gli uffici di statistica, vera rete di osservatorii della vita sociale, vengono ogni anno apprestando. Si agitarono vivamente, e anche più del bisogno, questioni secondarie (sulla definizione, sui limiti, sull'oggetto, ecc., della statistica), o trascendenti forse la competenza propria di questa disciplina (come la questione del *libero arbitrio*, che a nostro avviso appartiene alla filosofia); ma il maggior lavoro e il più fecondo fu speso a perfezionare le osservazioni per masse, e i metodi per la ricerca delle leggi empiriche. La separazione della statistica propriamente detta, considerata come metodo speciale di trattazione dei fenomeni collettivi, dalla demografia, come scienza quantitativa della popolazione, si avvia a diventare un fatto compiuto. La prima si afferma sempre meglio come un ramo nuovo e fecondo della logica; la seconda come una disciplina preparatoria ed ausiliare delle scienze sociali. Viceversa ha seguitato a perdere terreno la statistica nel vecchio significato della scuola di Conring e di Achenwall. L'indirizzo matematico, con felici risultati, impresso persino alla statistica biologica — indirizzo che ha forse per sè l'avvenire — ha fruttato novità interessanti, doverosamente nel corso di quest'opera menzionate. Ma la pleiade di scrittori che in un modo o nell'altro si son resi benemeriti dei progressi della statistica negli ultimi cinquant'anni, è tale e tanta che il fare alcuni nomi e tacerne alcuni altri parrebbe un riconoscere certa specie di servigi resi alla scienza e disconoscerne altre specie; la qual cosa è affatto lontana dalle nostre intenzioni.

ERRATA-CORRIGE

Pag.	Linea	Invece di	leggasi
29	penultima	seoond il Mill	secondo il Mill
31	11-12	quanto più estensivi	quanto più quelle sono estensive
46	4	giù grande	più grande
47	6	per 100 metri quadrati	per 100 metri quadrati,
74	penultima	se se ne ebbe una	se ne ebbe una
139	4 e seguenti	y 20 y 16,5 ecc.	$y = 20$ $y = 16{,}5$ ecc.
186	tra la 10ª e l'11ª	. . . .	Facendo $\log y = Y$ e $\log x = X$ si ha: $Y = 9.1854 - 1{,}45\,X$
191	penultima	$y = \frac{A}{x^{0,71507}} - 795$	$y = \frac{K}{x^{0,71507}} - 795$
»	ultima	dove con A	dove con K
206	20	cioè a 1886	cioè al 1886
»	36	fra 5,4 e 5,9	fra 5,40 e 5,49
213	3 del Sommario	Bernouilli	Bernoulli
216	terzultima	$^{18}/_{25}$	$^{18}/_{15}$
221	6 e 7	Gruppi dissimili $\begin{cases} L^s\, L^e \\ A^t\, A^e \end{cases}$	Gruppi dissimili $\begin{cases} L^t\, A^e \\ A^t\, L^e \end{cases}$
»	§ 5, linea 7	q_m	q^m
254	nota, linea 1	*Mesure de*	*Mesure des*